

DOCUMENT RESUME

ED 475 288

TM 033 722

AUTHOR McLean, James E., Ed.; Kaufman, Alan S., Ed.
 TITLE Research in the Schools, 2000.
 INSTITUTION Alabama Univ., Birmingham.; Mid-South Educational Research Association, MS.
 ISSN ISSN-1085-5300
 PUB DATE 2000-01-01
 NOTE 151p.
 PUB TYPE Collected Works - Serials (022)
 JOURNAL CIT Research in the Schools; v7 n1-2 Spr-Fall 2000
 EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS
 DESCRIPTORS *Academic Achievement; Art Education; Cooperative Learning; Corporal Punishment; Educational Finance; *Educational Research; Elementary School Students; Elementary Secondary Education; Expenditure per Student; High School Students; Intelligence Tests; Literacy; Magnet Schools; Prediction; Reliability; School Culture; School District Size; *Statistical Significance; *Teacher Attitudes
 IDENTIFIERS Chess

ABSTRACT

This journal publishes articles that focus on the result of applied educational research, scholarly reviews of research, descriptions of classroom techniques, studies of assessments, and other topics of interest to educational researchers. Issue number 1 of Volume 7 contains these articles: (1) "Predicting First Grade Achievement from Kindergarten Screening Measures: A Comparison of Child and Family Predictors" (Ronald K. Bramlett, R. Kevin Rowell, and Kristi Mandenberg); (2) "Academic Achievement and Magnet Schools" (Betty L. Dickson, Carolyn L. Pinchback, and Robert L. Kennedy); (3) "The Effects of Chess Instruction on the Mathematics Achievement of Southern, Rural, Black Secondary Students" (James P. Smith and Bob N. Cage); (4) "Changing Teacher Education Students' Attitudes toward Using Corporal Punishment in the Classroom" (Marlynn M. Griffin, Daniel H. Robinson, and Holli M. Carpenter); (5) "A Comparison of Attitudes and Expectations of Teachers and Peers towards Rural Gifted Female Adolescents" (Deborah Bretz, Neelam Kher, and Lorna J. Lacinda-Gifford); (6) "WISC-III Stability over a Three-Year Period in Students with Learning Disabilities" (Teresa Smith, Billy L. Smith, Ronald K. Bramlett, and Nancy Hicks); (7) "Cooperative Editing: An Effective Strategy for Error Reduction in Remedial Freshman Composition" (Sue S. Minchew and Neil G. Amos); and (8) "Factors Associated with Achievement in Educational Research Courses" (Anthony J. Onwuegbuzie, John R. Slate, Frances R. A. Paterson, Mary H. Watson, and Robert A. Schwartz). Issue number 2 contains, following an introduction, : (1) "Interpreting Statistical Significance Test Results: A Proposed New 'What If' Method" (Bruce Thompson and Kevin M. Kieffer); (2) "'First We Read great Book': Interfacing Literacy Instruction with the Arts in Two Urban Elementary Schools: Preservice Teachers' Case Quandaries and Accomplishments" (Janet C. Richards and Joan P. Gipe); (3) "A Critique of Methods Used To Answer the Question: Does money Make a Difference?" (Jim C. Fortune and Susan Spofford-Richardson); (4) "An Analysis of Factors Related to School-District Size in Arkansas" (Michael B. Gilbert); (5) "Adolescent Problems and Coping Strategies: Alternative School Versus Non-alternative School Students" (Deborah D. Vaughn, Ellen K. Slicker,

and Judith Van Hein); (6) "High-Stakes Accountability and Contextual Effects: An Empirical Study of the Fairness Issue" (Edward B. Reeves); (7) "Developing an Identity of Third Graders" (Sharon L. Newbill and Andrea Clements); and (8) "Measures of Work Culture in High and Low Performance Schools" (Darlene Y. Bruner and Bobbie Jean Greenlee). (SLD)

RESEARCH IN THE SCHOOLS

A nationally refereed journal sponsored by the
Mid-South Educational Research Association
and East Tennessee State University.

Volume 7, Numbers 1 and 2

2000

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL IN
MICROFICHE ONLY HAS BEEN GRANTED BY

J. McLean

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

20

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☐ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

729 033 722



RESEARCH IN THE SCHOOLS

A nationally refereed journal sponsored by the
Mid-South Educational Research Association
and the University of Alabama at Birmingham.

Volume 7, Number 1

Spring 2000

Predicting First Grade Achievement from Kindergarten Screening Measures: A Comparison of Child and Family Predictors	1
<i>Ronald K. Bramlett, R. Kevin Rowell, and Kristi Mandenberg</i>	
Academic Achievement and Magnet Schools	11
<i>Betty L. Dickson, Carolyn L. Pinchback, and Robert L. Kennedy</i>	
The Effects of Chess Instruction on the Mathematics Achievement of Southern, Rural, Black Secondary Students	19
<i>James P. Smith and Bob N. Cage</i>	
Changing Teacher Education Students' Attitudes toward Using Corporal Punishment in the Classroom	27
<i>Marlynn M. Griffin, Daniel H. Robinson, and Holli M. Carpenter</i>	
A Comparison of Attitudes and Expectations of Teachers and Peers Towards Rural Gifted Female Adolescents	31
<i>Deborah Bretz, Neelam Kher, and Lorna J. Lacinda-Gifford</i>	
WISC-III Stability Over a Three-Year Period in Students with Learning Disabilities	37
<i>Teresa Smith, Billy L. Smith, Ronald K. Bramlett, and Nancy Hicks</i>	
Cooperative Editing: An Effective Strategy for Error Reduction in Remedial Freshman Composition.	43
<i>Sue S. Minchew and Neil G. Amos</i>	
Factors Associated with Achievement in Educational Research Courses	53
<i>Anthony J. Onwuegbuzie, John R. Slate, Frances R. A. Paterson, Mary H. Watson, and Robert A. Schwartz</i>	

James E. McLean and Alan S. Kaufman, Editors

RESEARCH IN THE SCHOOLS

Information for Authors

Statement of Purpose

RESEARCH IN THE SCHOOLS (ISSN 1085-5300) publishes original contributions in the following areas: 1) *Research in Practice*--empirical studies focusing on the results of applied educational research including cross-cultural studies, 2) *Topical Articles*--scholarly reviews of research, perspectives on the use of research findings, theoretical articles, and related articles, 3) *Methods and Techniques*--descriptions of technology applications in the classroom, descriptions of innovative teaching strategies in research/measurement/statistics, evaluations of teaching methods, and similar articles of interest to instructors of research-oriented courses, 4) *Assessment*--empirical studies of norm-referenced, criterion-referenced, and informal tests in the areas of cognitive ability, academic achievement, personality, vocational interests, neuropsychological functioning, and the like, and 5) *Other* topics of interest to educational researchers. *RESEARCH IN THE SCHOOLS* is devoted to research conducted in *any* educational setting from a conventional elementary school or high school to a training program conducted within an industry. Likewise, there are no age restrictions on the sample, since the educational settings may include preschools, continuing education classes for adults, or adaptive skills courses in nursing homes. Studies conducted in settings such as clinics, hospitals, or prisons are ordinarily inappropriate for *RESEARCH IN THE SCHOOLS* unless they involve an educational program within such a setting. One goal of *RESEARCH IN THE SCHOOLS* is to provide a training ground for graduate students to learn effective reviewing techniques. Consequently, the journal utilizes a Graduate Student Editorial Board composed mostly of students in educational psychology and educational research. Members of this Editorial Board, each sponsored by a professor, provide supplementary reviews for a selection of submitted articles, and receive both direct and indirect feedback of the quality of these reviews.

Preparing Manuscripts

Authors should prepare manuscripts in accordance with the stylistic rules and guidelines delineated in the *Publications Manual of the American Psychological Association* (4th ed., 1994), which is available from: Order Department, American Psychological Association, PO Box 2710, Hyattsville, MD 20784. Number the pages consecutively. All manuscripts will be subject to editing for sexist language.

Author Identification

Authors should put the complete title of the article on the first text page, but they should exclude their names. Subsequent pages should include only a running head. They should prepare a separate sheet with the complete title of the article and their names and affiliations; this procedure will ensure anonymity in the review process. Authors should supply addresses and phone numbers, and electronic mail addresses and fax numbers (if available), for potential use by the editorial staff and, later, by the production staff. Unless otherwise stated, the first-named author will be sent correspondence, galley proofs, copyright forms, and so forth.

Submission of Manuscripts

Submit manuscripts in triplicate to James E. McLean, Co-Editor, *RESEARCH IN THE SCHOOLS*, School of Education, 233 Educ. Bldg., The University of Alabama at Birmingham, 901 13th Street, South, Birmingham, AL 35294-1250. Please direct questions to jmclean@uab.edu. All copies should be clear and readable; dot matrix is acceptable only if it meets these qualities of legibility. Length of the manuscripts, including references and tables, should ordinarily range from about 10 to 40 typed, double-spaced, 8-1/2 X 11-inch pages, using 11-12 point type. Abstracts are limited to 125 words. Brief reports of research are not encouraged. Authors are encouraged to keep a hard copy of the manuscript to guard against loss. It is assumed that all manuscripts submitted for publication are original material and have not been simultaneously submitted for publication elsewhere. When manuscripts are accepted for publication, authors are encouraged to submit the final version on a computer disk along with the hard copy.

Copyright and Permissions

Authors are granted permission to reproduce their own articles for personal use. Others must request permission to reproduce tables, figures, or more than 500 words of text from the editors. Copyright © 2000 by the Mid-South Educational Research Association.

Volume 7, Number 1

Spring 2000

EDITORS

James E. McLean, *University of Alabama at Birmingham*
Alan S. Kaufman, *Yale University School of Medicine*

PRODUCTION EDITOR

Margaret L. Rice, *The University of Alabama*

COPY EDITOR

Michele G. Jarrell, *University of Alabama at Birmingham*

EDITORIAL BOARD

Jan E. Alm, *Uppsala University (Sweden)*
Charles M. Achilles, *Eastern Michigan University*
Margery Arnold, *Trinity University*
Mark Baron, *University of South Dakota*
Alan W. Brue, *Carrollton City (Georgia) Schools*
Robin Cook, *Wichita State University*
Larry G. Daniel, *University of North Texas*
James M. Ernest, *State University of New York at Buffalo*
Daniel Fasko, Jr., *Morehead State University*
Glennelle Halpin, *Auburn University*
Harrison Kane, *University of Nevada, Las Vegas*
JinGyu Kim, *National Institute of Professional Administration (Korea)*
Jwa K. Kim, *Middle Tennessee State University*
Frank R. Lawrence, *University of Alabama at Birmingham*
Robert E. Lockwood, *Alabama State Department of Education*
Robert Marsh, *Chattanooga State Tech. Community College*
Jerry Mathews, *Auburn University*
Charles McLafferty, Jr., *University of Alabama at Birmingham*
Peter C. Melchers, *University of Cologne (Germany)*
Claire Meljac, *Hospital Sainte-Anne (France)*
Soo-Back Moon, *Catholic University of Hyosung (Korea)*
David T. Morse, *Mississippi State University*
Jack A. Naglieri, *Ohio State University*
Anthony J. Onwuegbuzie, *Valdosta State University*
Cecil R. Reynolds, *Texas A & M University*
Janet C. Richards, *The University of Southern Mississippi*
Michael D. Richardson, *Georgia Southern University*
Dan Robinson, *University of Texas at Austin*
John R. Slate, *University of Texas at El Paso*
Paula Stanley, *Radford University*
Bruce Thompson, *Texas A & M University*

GRADUATE STUDENT EDITORIAL BOARD

Keri L. Altig, *University of Nevada, Las Vegas*
Adam S. Bristol, *Yale University*
James M. Graham, *Texas A & M University*
Debbie Hahs, *The University of Alabama*
Keith Hyatt, *University of Nevada, Las Vegas*
James C. Kaufman, *Yale University*
Kevin M. Kieffer, *Texas A & M University*
Ginny Lane, *University of North Texas*
Lynn C. Minor, *Valdosta State University*
William T. Pepper, *Mississippi State University*
Catherine A. Stacy, *University of Texas at Austin*

Predicting First Grade Achievement from Kindergarten Screening Measures: A Comparison of Child and Family Predictors

Ronald K. Bramlett, R. Kevin Rowell, and Kristi Mandenberg

The University of Central Arkansas

Kindergarten screening is a common practice in the United States. Educators and mental health professionals are interested in finding methods that will lead to effective interventions for children and families. This study examined three kindergarten screening methods. The Parenting Stress Index, the Behavior Assessment System for Children, and an informal letter/number probe were compared in predicting first grade reading and math. Correlational analyses showed that each method was significantly related to first grade reading and math, as measured by the Woodcock-Johnson Psychoeducational Battery - Revised. Multiple regression analyses revealed that letter and number recognition in kindergarten were the best predictors of first grade reading and math. Child behavior and parenting stress were also significantly related to reading and math, although the additional variance explained was relatively small. Discriminant analyses revealed that the measures were good at identifying those children who were not "at-risk" but the overall prediction of adequate versus poor readers was less than chance. All of the methods produced large percentages of false positives and negatives. Implications for screening efforts are discussed.

The practice of kindergarten screening to identify children with potential academic and social difficulties has become a common practice in the United States in the past two decades (Agostin & Bain, 1997; Cannella & Reiff, 1989). Educators and mental health professionals (e.g., psychologists, counselors, social workers, etc.) are interested in utilizing assessment procedures that are most likely to predict school success and lead to appropriate intervention strategies. Typically, the focus of kindergarten screening has been in domains such as language, cognitive abilities, motor skills (gross and fine), social-emotional functioning and self-help (e.g., McLoughlin & Rausch, 1990). While these areas have been identified as important for some intervention efforts (Lichtenstein & Ireton, 1991), developmental screening measures often lack the technical rigor to make accurate predictions or placements (Shepard, 1997). In fact, depending on the instrument that is selected, very different decisions can be made related to which children are identified, as well as targets for intervention (e.g., Hall, Bramlett, Barnett, & Cox, 1994).

Ronald K. Bramlett is an Associate Professor in the Department of Psychology and Counseling at the University of Central Arkansas. R. Kevin Rowell is Associate Professor of Psychology and Counseling, University of Central Arkansas. Kristi Mandenberg is in the Department of Pediatrics, University of Arkansas for Medical Sciences. This study was funded through the Head Start-Public School Early Transition Project. Please address correspondence to Ronald K. Bramlett, Department of Psychology and Counseling, Box 4915, University of Central Arkansas, Conway, AR 72035 or via e-mail at Ronkb@mail.uca.edu.

Parent and family variables are not usually a focal point of kindergarten screening efforts. However, the families of young children "at-risk" for school problems often have critical needs that merit intervention consideration (Abidin, 1990b; Bramlett, Hall, Barnett, & Rowell, 1995). Parenting stress has been identified as an important family variable for reasons including its effects on the parent-child relationship (Laosa & Sigel, 1982; Rutter, 1985; Siegel, McGillicuddy-DeLesi, & Goodnow, 1992) and its prominence among the problems for which parents seek assistance. Abidin (1990a) defined parenting stress as the relative magnitude of stress in the parent-child system. Parenting stress may be disruptive to families (Patterson, 1988), and the potential risks for negative child outcomes (both academic and behavioral) has been documented.

The relationship between parenting stress and behavior problems in children is reciprocal. Child temperament, stressful life events, and parent characteristics (e.g., temperament, marital relationship) qualitatively and quantitatively impact the type of parent-child interaction and influence the amount of stress experienced (Abidin, 1990a). Mouton and Tuma (1988) reported a consensus in the research suggesting that maternal variables affect child variables and vice versa. A study conducted by Bendell, Stone, Field, and Goldstein (1989) also showed the circular relationship between parenting stress and child outcomes. These authors found that children's spelling achievement, self-esteem, and conduct significantly contributed to parenting stress as reported by mothers in their sample.

Parenting stress has been found to be related to a wide range of child behaviors. The association between

disruptive behaviors and maternal stress is illustrated in Eyberg, Boggs, and Rodriguez (1992), who found that parenting stress is strongly related to disruptive behaviors. That is, parents who report disruptive behavior by their children report higher levels of stress. Children with difficult temperaments are more likely to elicit coercive parenting strategies than compliant children because of the stress that is created (Anderson, Lytton, & Romney, 1986).

Cowan, Cowan, Schulz, and Hemming (1993) found that poor marital relationships and inadequate parenting (identified as two important characteristics of parenting stress) were predictive of kindergarten achievement. It appears that a mother's reported experience of life stress is correlated with her child's academic achievement as early as kindergarten thereby reinforcing the possible need to target parenting stress in early intervention (Pianta, Egeland, & Sroufe, 1990).

Fuller and Rankin (1994) found that parents with emotionally disturbed children (those with a profile similar to a diagnosis of ADD) experienced higher levels of parenting stress than parents of learning disabled children or those in regular education. Other studies have found higher levels of parental stress in families with ADHD children compared to normal controls (Anastopoulos, Guevremont, Shelton, & DuPaul, 1992). Breen and Barkley (1988) found a relationship between child psychopathology (particularly disorders with externalizing symptoms such as ADHD) and parenting stress. Additionally, preliminary findings by Cuccaro, Holmes, and Wright (1993) indicated that parents of preschoolers with externalizing problems experienced greater stress than parents with preschoolers exhibiting internalizing symptoms. Thus, impaired academic achievement, dysfunctional interactions, and child behavior problems have been identified as correlates of parenting stress.

In addition to parenting stress, pre-academic skills are valid predictors of success. For example, letter and number naming ability has been shown to be one important early literacy skill, and knowledge of letters at the beginning of kindergarten has been found to be related strongly to later reading and math achievement (Dermott, Pinzari, Dodd, & McIntire, 1980). For early intervention purposes, letter naming fluency (rate of correct letters read per minute) may be an important pre-reading skill to assess and monitor (Kaminski & Good, 1996). It is important to compare these pre-academic skills to other child and family variables.

While parenting stress and child behavior are identified as correlates of school success, the relative importance of assessing parenting stress for predicting later school success needs closer examination. For example, does knowledge of parenting stress predict which children are more likely to be at-risk for school problems? And

how can this information be used to assist families in developing early interventions? Does parenting stress add anything to the prediction of first grade reading and math achievement beyond direct assessment of academic skills?

Purpose of the Study

The purpose of this study was to examine the predictability of parenting stress and child behaviors (problematic and adaptive) in kindergarten or first grade reading and math in a population of former Head Start children and their families. In addition, the study examined the validity of parenting stress, child behavior, and kindergarten entry-level academic skills in predicting potential "at-risk" students for reading and math.

Method

Participants

Parents or guardians of kindergarten children in nine rural public schools completed the PSI (Abidin, 1990a) and the BASC (Reynolds & Kamphaus, 1992) in the fall of their children's kindergarten year ($N = 92$). All of the children attended a Head Start center the previous year and were part of a larger study examining the effects of transition services for Head Start children. Head Start children are important to study because they may be particularly vulnerable to environmental and familial stress and in need of interventions aimed at key family variables such as parenting stress (Mantzicopoulos, 1997; Teo, Carlson, Mathieu, Egeland, & Sroufe, 1996). Achievement measures were administered to the children in the fall of kindergarten and the spring of first grade. The child sample was composed of 47% males and 53% females, 100% were white/Caucasian, and the mean age was 5 years, 6 months (range 5 years, 0 months - 5 years, 11 months). The parents' mean age was 29 years, 6 months. The respondents included 92% mothers, 5% guardians, and 3% fathers. The median family income was approximately \$1,500 per month. Seventy-eight percent of the parents had at least the equivalent of a high school diploma.

Materials

Parenting Stress Index. The Parenting Stress Index (PSI; Abidin, 1990a) is a self-report instrument used to identify parent-child interactions, which may indicate a risk for the development of maladaptive behaviors. The PSI consists of 101 items and yields three domain scores: the Child Domain, the Parent Domain, and an optional Life Stress Index Scale. The Child Domain measures child characteristics such as adaptability, mood, demand-ingness and activity level. The Parent Domain measures parent characteristics such as depression, sense of parental competence, investment in parenting, and spousal support.

PREDICTING FIRST GRADE ACHIEVEMENT

A Total Score is obtained by summing the Child and Parent Domains. High total scores (>260) reveal the possibility of high stress and should be considered for possible intervention. The normative group was primarily based on responses by 534 parents attending pediatric clinics in central Virginia.

The alpha reliabilities range from .62 to .70 for the subscales of the Child Domain, and from .55 to .80 for the Parent Domain subscales. For the domains, the reliabilities are .89 (Child Domain), .93 (Parent Domain), and .95 (Total Stress). Overall, the test-retest reliability estimates range from a high of .96 for the Total Stress Score, .70 for the Parent Domain, and .55 for the Child Domain ($n = 37$, 1 year interval).

Content validity was based on a comprehensive literature review and professional ratings of item relevance. Several studies are reported in the manual that support the concurrent and construct validity to varying degrees and in complex patterns of findings (Abidin, 1990a). The PSI was selected because it has been utilized as a criterion measure in studies of stressful parent-child interactions (Abidin, 1990a).

Behavior Assessment System for Children. The Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992) is a multi-measure instrument designed to provide a holistic view of the child and problem by obtaining information from several sources (parent, child, teacher, observation, and historical records). The BASC was developed for use in the diagnosis, treatment, and educational classification of children with behavioral and emotional disorders. The Parent Rating Scale (PRS) for ages 4-5 was utilized in this study and consists of 131 items to be completed by a parent or guardian. The parent indicates the frequency of the child's behavior over the last six months by responding *never*, *sometimes*, *often*, or *almost always*. The PRS yields *T* scores and percentiles for clinical scales in three composite areas: Externalizing Problems (i.e., Hyperactivity, Aggression, Conduct Problems), Internalizing Problems (i.e., Anxiety, Depression, Somatization), and Adaptive Skills (i.e., Adaptability, Social Skills). A Behavioral Symptoms Index is a combination of the central scales of the clinical composites (Externalizing and Internalizing Problems). The Adaptive Skills composite summarizes prosocial adaptive and social skills. The General Norms sample consisted of 309 parents of children representing the general population tested in 116 sites in the United States. The Clinical Norms sample consisted of 100 parents of children identified by schools or clinical settings as having behavioral or emotional problems.

Internal consistency reliabilities for the composite scores are .80 to .90 and .88 to .94 for the Behavior

Symptoms Index. The PRS has a median test-retest reliability of .85. Construct validity of the PRS is illustrated by correlations of .65 to .84 with the Child Behavior Checklist (CBCL; Achenbach, 1991). Content validity was established using covariance structure analysis and principal-axis factor analysis of the items.

Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R). The Reading and Mathematics clusters and subtests (Word Identification, Passage Comprehension, Calculation, and Applied Problems) of the WJ-R (Woodcock & Johnson, 1990) were examined in this study. The WJ-R was standardized nationally on 6,359 subjects. The Reading cluster is a combination of Letter-Word Identification and Passage Comprehension subtests. Its median reliability is .94 in the kindergarten to grade 12 range. The Mathematics cluster contains Calculation and Applied Problems subtests. The median reliability is .93 in kindergarten to grade 12 range. The WJ-R was used in this study because it was one of the more commonly used achievement tests.

Letter/Number Probes. Two sets of letter/number probes were constructed that included the 26 letters of the alphabet (upper and lower case, in random order) and numbers (1-20, in random order). The letters and numbers were typed on single 8.5 x 11 inch, white paper. Children were asked to read as many letters as they could in a one-minute time period. The two reading probes were added together and divided by 2 in order to get a score. The same procedure was followed for number

Procedures

The PSI and BASC were completed in counter-balanced order by the parents or guardians in the fall of kindergarten. The letter/number probes were administered to the children in the fall of kindergarten, and the WJ-R was administered in the spring of first grade. The probes and WJ-R were administered by trained data collectors. Each data collector demonstrated competency with the procedures prior to administration. Each data collector attended training sessions where the procedures were modeled to them; they were observed by the trainer in administering the procedures; corrective feedback was given when necessary; and finally, all protocols were checked for accuracy by a data collection supervisor.

Data Analysis

Two separate multiple regression analyses were conducted to determine the degree to which reading and math achievement in first grade could be predicted from children's behavior, adaptive functioning, parenting stress and letter/number recognition in kindergarten. In the first analysis, the two primary composite scores of the BASC,

Behavior Symptom Index (BSI) and Adaptive Skills (ADAP), and the PSI total scores were used as predictors of the Woodcock-Johnson reading and math scores. In the second analysis, letter/number recognition was added to the model to determine what it contributed to the prediction of reading and math scores above the BASC and PSI scores.

As a secondary analysis, discriminant analyses were conducted to determine the ability of the measures in predicting first grade reading and math classification. Students were divided into "adequate" or "poor" categories based on their first grade performance in reading and math on the WJ-R. Students with scores on the WJ-R at or below a standard score of 85 were judged to be "poor." "Adequate" performers were those with scores of 86 and above. This criterion was used because it represents one standard deviation below the mean. While this is not meant to be a perfect criterion, approximately 15% of the student population falls into this category and are generally more likely to experience reading and math problems.

Results

Descriptive statistics for the PSI, BASC, letter/number probes, and WJ-R scores are presented in Table 1. The means and standard deviations were very similar to the normative samples on these measures. On the letter/number probes, there was wide variability of scores (range 0 - 46; 0 - 51). Eleven percent of the children had raw scores of zero on letters, and 6% received raw scores of zero on numbers. Due to the lack of normality in the letter/number probe data, a square root transformation was applied before analyses were conducted.

Table 1
Descriptive Statistics for the PSI, BASC, Letter/Number Probes in Kindergarten and the WJ-R (Reading and Math) in First Grade ($N = 92$)

Variable	Mean	SD	Range
PSI	230	38	124 - 347
BASC			
Adaptive Scale	48	9	19 - 75
Behavioral Symptom Index	54	10	23 - 104
Letters *	10	10	0 - 46
Numbers *	12	12	0 - 51
WJ-R Reading	101	17	56 - 141
WJ-R Math	104	18	44 - 159

* Values before square root transformation

Correlations were computed between all of the variables. The highest correlations were between the Letter and Number probes ($r = .82$) with the lowest occurring between Adaptive Skills (ADAP) and Number (square root) probes ($r = .04$). Correlations were in logical

directions. For example, parenting stress and their children's adaptive skills were negatively correlated ($r = -.47$, $p < .01$) whereas parenting stress and behavioral problems were positively correlated ($r = .32$, $p < .01$) (Table 2).

Table 2
Correlations Between PSI, BASC (BSI and ADAP), and Letter/Number Probes in Kindergarten and WJ-R Reading and Math in First Grade ($N = 92$).

Variable	1	2	3	4	5	6	7
1. PSI	--	.32**	-.47**	-.10	-.09	-.20	-.34**
2. BSI		--	-.38**	-.23*	-.15	-.18	-.17
3. ADAP			--	.05	.04	.31**	.27*
4. Letters Sqrt				--	.82**	.43**	.39**
5. Numbers Sqrt					--	.38**	.41**
6. WJ-R Reading						--	.72**
7. WJ-R Math							--

* $p \leq .05$ ** $p \leq .01$

Multiple Regression Analyses. In Model 1, the PSI total score was entered first because its partial correlation with reading scores was greater than that of BASC scores (Table 3). The regression model was not significant, $F(1,90) = 3.78$, $p = .06$, and only accounted for about 3% of the variance in reading scores. Even with the addition of BASC Behavior Symptoms Index (BSI) scores in Model 2, the regression model was not significant, $F(2,89) = 2.62$, $p = .08$, and less than 1% of predicted variance was added. Results of these two models indicated that parents' ratings of children's problematic behaviors and family stress did not provide valid prediction of reading achievement in the first grade.

In Model 3, the BASC Adaptive Skills (ADAP) was added after PSI and BASC (BSI) to determine if adaptive functioning was related to reading scores. In this model, ADAP was a statistically significant predictor, $\beta = .26$, $t = 2.15$, $p = .03$, and the overall model was also statistically significant; however, only 7.2% of variance in reading scores was explained.

Model 4 was constructed by adding kindergarten letter recognition scores into the model then adding PSI, BSI, and ADAP scores in order to determine if parents' ratings of children's behavior, family stress, and adaptive functioning adds to the prediction of reading scores over and above simply using letter recognition. Because Letter and Number recognition were highly correlated, $r = .82$, $p < .001$, and because letter recognition would be more appropriate than number recognition as a predictor of reading, only letter recognition was used in constructing this model. As seen in Table 3, the model was statistically significant, $F(4,87) = 7.98$; $p < .001$, and accounted for 23.5% of the variance in reading scores. Letter recognition, $\beta = .42$, $t = 4.44$, $p < .001$, accounted for the vast

PREDICTING FIRST GRADE ACHIEVEMENT

majority of variance followed by ADAP, $\beta = .28$, $t = 2.62$, $p = .01$. PSI and BSI were not significant predictors.

Table 3
Building the Regression Model Using Kindergarten
Variables to Predict WJ-R Reading Score
in First Grade ($N = 92$).

Variable	Model 1	Model 2	Model 3	Model 4
Predictors				
PSI	-.20	-.16	-.13	-.04
BSI		-.13	-.03	.04
ADAP			.26*	.28**
Letter Sqrt				.42***
Model				
F	3.78	2.62	3.63	7.98
df	1.90	2.89	3.88	4.87
p_{calc}	.06	.08	.022	.000
Adj. R^2	.030	.034	.072	.235

Note. Regression coefficients are β weights.

* $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$

To construct the most parsimonious model, only letter recognition and adaptive functioning were used as predictors in the final model (Table 4). Again, letter recognition accounted for the majority of the variance, $\beta = .42$, $t = 4.56$, $p < .001$, followed by adaptive functioning, $\beta = .29$, $t = 3.15$, $p = .002$. This final model was statistically significant, $F(2,89) = 16.20$; $p < .001$, and accounted for 25% of the variance in reading scores. Analysis of the residuals through a probability plot and by plotting the residuals against predicted reading values indicated that the assumptions of the model were not violated.

Table 4
Final Regression Model Using Kindergarten Variables to
Predict WJ-R Reading Scores in First Grade ($N = 92$).

Variable	B	SE B	β	p_{calc}
Constant	65.45	8.21	0.00	0.000
Letter Sqrt	4.08	0.89	0.42	0.000
ADAP	0.52	0.17	0.29	0.002

$F = 16.20$, $df = 2,89$, $p_{calc} < .001$, Adj. $R^2 = .250$

Similar to the prediction of reading scores, multiple regression analyses were conducted to determine if PSI, BSI, and ADAP scores predicted Woodcock-Johnson math achievement. In Model 1, PSI was again entered first because, like reading scores, its partial correlation with math scores was greater than that of BSI scores (Table 5). The regression model was statistically significant, $F(1,90) = 11.57$, $p < .001$, accounting for about 10% of the variance in math scores, and PSI was statistically significant, $\beta = -.34$, $t = -3.40$, $p < .001$. With

the addition of BSI scores in Model 2, the regression model was again significant, $F(2,89) = 5.96$, $p = .004$; however, the BSI scores did not add any predictive value, and in fact, reduced the standardized coefficient value of PSI somewhat ($\beta = -.32$). The amount of predicted variance declined slightly to 9.8%. The results, therefore, of these two models indicated that family stress was slightly predictive of math achievement but ratings of children's behavior were not a significant predictor.

Table 5
Building the Regression Model Using Kindergarten Variables
to Predict WJ-R Math Scores in First Grade ($N = 92$).

Variable	Model 1	Model 2	Model 3	Model 4
Predictors				
PSI	-.34***	-.32**	-.26**	-.24*
BASC		-.07	-.03	.02
ADAP			.14	.15
Number Sqrt			.38***	
Model				
F	11.57	5.96	4.44	8.29
df	1.90	2.89	3.88	4.87
p_{calc}	.001	.004	.006	.000
Adj. R^2	.104	.098	.102	.243

Note. Regression coefficients are β weights.

* $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$

In Model 3, ADAP was added after PSI and BSI to determine if adaptive functioning was related to math scores (Table 5). In this model, the PSI remained statistically significant as a predictor, $\beta = -.26$, $t = -2.30$, $p = .02$; however, neither BSI or ADAP scores were significant predictors. The overall model was also statistically significant with 10.2% of the variance in math scores explained.

Model 4 was constructed by adding kindergarten number recognition scores into the model then adding PSI, BSI, and ADAP scores in order to determine if ratings of children's behavior, family stress, and adaptive functioning add to the prediction of math scores over and above simply using number recognition. As discussed above, Letter and Number recognition were highly correlated, $r = .82$, $p < .001$; therefore number recognition would be more appropriate than letter recognition as a predictor of math. As seen in Table 5, the model was statistically significant, $F(4,87) = 8.29$; $p < .001$, and accounted for 24.3% of the variance in math scores. Number recognition, $\beta = .38$, $t = 4.16$, $p < .001$, accounted for the vast majority of variance followed by PSI, $\beta = -.24$, $t = -2.28$, $p = .02$. ADAP and BSI were not significant predictors.

Similar to the reading model, only the best predictors were used to obtain a final regression model that would be parsimonious. For the math model, only number recognition and parenting stress were used as predictors (Table 6). As expected, number recognition accounted for the majority of the variance, $\beta = .38$, $t = 4.17$, $p < .001$, followed by parenting stress, $\beta = -.30$, $t = -3.32$, $p = .001$. This final model was statistically significant, $F(2,89) = 15.52$; $p < .001$, and accounted for 24.2% of the variance in math scores. Similar to the reading model, analysis of the residuals through a probability plot and by plotting the residuals against predicted math values indicated that the assumptions of the model were not violated.

Table 6
Final Regression Model Using Kindergarten Variables
to Predict WJ-R Math Scores in First Grade ($N = 92$).

Variable	<i>B</i>	<i>SE B</i>	β	<i>p</i> _{calc}
Constant	124.90	11.10	0.00	.000
Number	4.12	0.99	0.38	0.000
PSI	-0.15	0.04	-0.30	0.003

$F = 15.52$, $df = 2,89$, $p_{calc} = 0.000$, $Adj. R^2 = 0.242$

Discriminant Analyses. A predictive discriminant analysis was conducted secondarily to determine the classification accuracy of the predictors generated in the reading and math regression models. Students were considered to be at-risk in reading or math if their first-grade Woodcock-Johnson respective scores were below 85 (-1 s.d.). In order to maintain the assumption of normality for variables in a discriminant analysis, the Letter and Number recognition scores were converted into more normal distributions by using a square root transformation. The correlations of these square root values with their respective Woodcock-Johnson reading and math scores were similar to the correlations obtained by their raw scores.

Table 7 shows the classification accuracy of ADAP and Letter recognition in predicting at-risk readers in the first grade. When entered separately into the discriminant analysis, ADAP was statistically significant, $F_{1,90} = 5.02$, $p = .028$, for the Wilks' lambda value and had a better overall accuracy (68%) than Letter recognition square root (57%), which was also statistically significant, $F_{1,90} = 7.39$, $p = .008$. The combination of both variables in predicting reading scores was statistically significant as well, $F_{2,89} = 6.20$, $p = .003$, but was no more accurate (68%) than ADAP alone. In this sample, however, the proportion of those who could be classified by chance was 72.6%, that is, lumping all cases into the larger category of "not at-risk". As a result, the discrimination value of the model of classifying both "at-risk" and "not at-risk" was actually less than chance.

Table 7
Classification Accuracy of Kindergarten Measures in Predicting
At-Risk Reading Scores on the Woodcock-Johnson
in First Grade ($N = 92$).

Variable	True +	True -	False +	False -	<i>F</i> [*]	Hit Rate	PPV	NPV
ADAP	10	53	24	5	.028	.68	.29	.91
Letter Sqrt	10	46	31	5	.008	.61	.24	.90
Letter Sqrt + ADAP	11	52	25	4	.003	.68	.31	.93

PPV = Positive Predictive Value NPV = Negative Predictive Value

*Note: *F* is the Wilks' lambda estimate

Positive predictive value (PPV), which is the ratio of those who were actually "at-risk" based on WJ-R scores (True positive) to all those predicted to be "at-risk" by the model (True positive + False positive), ranged from .24 for Letter recognition square root to .31 for the combined model. Negative predictive value (NPV), which is the ratio of those who actually were not "at-risk" (True negative) to all those predicted not to be "at-risk" by the model (True negative + False negative), ranged from .90 for ADAP alone to .93 for Letter recognition. Basically, the model was much better at identifying those who were not at-risk than identifying those who were at risk; however, its utility as a screening instrument for children at risk may still be of value. The relatively high number of false positives and very low number of false negatives would be beneficial in educational settings where further testing of those at-risk is feasible.

Table 8 shows the classification accuracy of PSI and Number recognition in predicting those who are at-risk for math in the first grade. PSI was statistically significant, $F_{1,90} = 4.12$, $p = .045$, for the Wilks' lambda value and had a better overall accuracy (68%) than Number recognition square root (61%), which was not statistically significant, $F_{1,90} = 2.35$, $p = .129$. The combination of both variables in predicting math scores was statistically significant as well, $F_{2,89} = 3.04$, $p = .053$, but was only slightly more accurate (67%) than PSI alone. In this sample, however, the proportion of those who could be classified by chance was 77.3%, that is, lumping all cases into the larger category of "not at-risk." As a result, the discrimination value of the model of classifying both "at-risk" and "not at-risk" was actually less than chance.

Positive predictive values (PPV) ranged from .21 for Number recognition square root to .26 for PSI as the only model predictor. Negative predictive values (NPV) ranged from .94 for Number recognition square root to .96 for PSI alone. Like the model for reading, the math model was much better at identifying those who were not at-risk than identifying those who were at-risk.

PREDICTING FIRST GRADE ACHIEVEMENT

Table 8
Classification Accuracy of Kindergarten Measures in Predicting
At-Risk Math Scores on the Woodcock-Johnson
in First Grade ($N = 92$).

Variable	True +	True -	False +	False -	F*	Hit Rate	PPV	NPV
PSI	10	51	29	2	.045	.66	.26	.96
Number Sqrt	9	46	34	3	.129	.60	.21	.94
PSI + Number Sqrt	9	53	27	3	.053	.67	.25	.95

PPV = Positive Predictive Value NPV = Negative Predictive Value

* Note: F is the Wilks' lambda estimate

Discussion

Educators and mental health professionals are interested in identifying the best predictors of school success. Most kindergarten screening efforts focus on child variables such as developmental readiness. However, it may be essential to examine family variables that affect important child outcomes. This study examined three potential kindergarten screening measures and their ability to predict first grade reading and math performance. The Parenting Stress Index, the Behavior Assessment System for Children, and an informal measure of pre-kindergarten academic skills (letter/number fluency) were compared to each other and utilized to predict first grade scores on the reading and math subtests of the Woodcock-Johnson-Revised.

The results of this study showed that the child's fluency in recognizing letters and numbers at the beginning of kindergarten was the best predictor of reading and math compared to parenting stress and parental ratings of problematic behaviors and adaptive skills. While the study found significant correlations between parenting stress, child behavior, and achievement, these variables did not explain additional variance of the first grade achievement scores above the variance explained by letter/number fluency. The classification accuracy of those predictors generated in the regression models was somewhat mixed. The predictors were good at identifying those children who were not "at-risk," but the overall accuracy was less than chance. However, they may still have utility for screening.

Implications for Practice

Kindergarten screening is often used to identify children who are likely to need some type of intervention. In these cases, dichotomous decisions related to who needs additional help or services are made. Typically the

purpose of screening is to identify those who are most likely to need extra assistance (True positives). As would be expected based on previous research, the letter and number probes were significant predictors of reading and math (e.g., Dermott et al., 1980; Kaminski & Good, 1996; Kelly & Peverly, 1992). Parenting stress and parents' ratings of behaviors add little explanation of variance above these pre-academic skills to the overall prediction of reading and math. While the letters and numbers tasks over-identified children who ultimately were not "at-risk" (False positives), these measures were good at identifying the true positives and true negatives. In screening, educators can probably justify over-identifying children in order to make sure they identify those most at-risk.

The letters/numbers method is probably worth considering as a screening tool because it is both time- and cost-efficient and links directly to what is taught in the classroom. Letter and number probes can be used frequently and are a good way to measure the child's progress within the curriculum. Compared to the other methods in this study, it is probably a better fit with the actual school curriculum. Teachers could use this method to focus tutoring efforts on children who struggle with these tasks.

Adaptive skills was also shown to be a moderate predictor of achievement similar to other studies (e.g., Agostin & Bain, 1997). Social skills is probably an important factor to focus on in kindergarten for screening and intervention. Certainly, children who have the skills to elicit positive behaviors from their peers and teachers are more likely to succeed in school situations. However, in this study, adaptive skills added little to the prediction of reading and math above letters and numbers.

While parenting stress is correlated with achievement, it is more difficult for schools to intervene directly in this area. Parenting stress is an important variable to consider especially if interventions are going to include parents as change agents. Based on the results of this study, it does not appear that wide-scale screening of parenting stress is warranted. However, for individual cases, parenting stress might be important to assess and intervene. Referrals to mental health services may be warranted in cases of high parental stress.

This study was limited by a homogeneous, white sample. Other racial and ethnic groups should be examined. The current results did not support the exclusive use of any particular method in determining "school readiness" for kindergarten children. The significant numbers of false decisions that could be made from these data should deter educators from using any of these methods to assign children to groups (e.g., developmental kindergarten, etc). However, if follow-up assessment with "at-risk" children

is possible, the large numbers of false decisions might be significantly reduced. The results of this study may provide additional information on utilizing child and parenting measures in guiding intervention efforts.

References

- Abidin, R. R. (1990a). *Parenting Stress Index*. Odessa, FL: Psychological Assessment Resources.
- Abidin, R. R. (Ed.) (1990b). Special issue on the stress of parenting. *Journal of Clinical Child Psychology*, 19, 298-391.
- Achenbach, T. M. (1991). *Child Behavior Checklist*. Burlington, VT: Author.
- Agostin, T. M., & Bain, S. K. (1997). Predicting early school success with developmental and social skills screeners. *Psychology in the Schools*, 34, 219-228.
- Anastopoulos, A. D., Guevremont, D. C., Shelton, T. L., & DuPaul, G. J. (1992). Parenting stress among families of children with attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, 20, 503-520.
- Anderson, K. E., Lytton, H., & Romney, D. M. (1986). Mothers' interactions with normal and conduct-disordered boys: Who affects whom? *Developmental Psychology*, 22, 604-609.
- Bendell, R. D., Stone, W., Field, T., & Goldstein, S. (1989). Children's effects on parenting stress in a low income, minority population. *Topics in Early Childhood Special Education*, 8, 58-71.
- Bramlett, R. K., Hall, J. D., Barnett, D. W., & Rowell, R. K. (1995). Child developmental/educational status in kindergarten and family coping as predictors of parenting stress: Issues for parent consultation. *Journal of Psychoeducational Assessment*, 13, 157-166.
- Breen, M. J., & Barkley, R. A. (1988). Child psychopathology and parenting stress in girls and boys having attention deficit disorder with hyperactivity. *Journal of Pediatric Psychology*, 13, 265-280.
- Cannella, G. S., & Reiff, J. C. (1989). Mandating early childhood entrance/retention assessment: Practices in the United States. *Child Study Journal*, 19, 83-99.
- Cowan, P. A., Cowan, C. P., Schulz, M. S., & Heming, G. (1993). Prebirth to preschool family factors in children's adaptation to kindergarten. In R. Parke & S. Kellam, (Eds.), *Exploring family relationships with other social contexts. Advances in Family Research* (Vol. 4). Hillsdale, NJ: Erlbaum.
- Cuccaro, M. L., Holmes, G. R., & Wright, H. H. (1993). Behavior problems in preschool children: A pilot study. *Psychological Reports*, 72, 121-122.
- Dermott, R. A., Pinzari, S., Dodd, E. R., & McIntire, W. (1980). Number knowledge for prediction of specific reading skills and achievement in first graders. *Reading Improvement*, 17, 86-93.
- Eyberg, S. M., Boggs, S. R., & Rodriguez, C. M. (1992). Relationships between maternal parenting stress and child disruptive behavior. *Child and Family Behavior Therapy*, 14, 1-9.
- Fuller, G. B., & Rankin, R. E. (1994). Differences in levels of parental stress among mothers of learning disabled, emotionally impaired, and regular school children. *Perceptual and Motor Skills*, 78, 583-592.
- Hall, J. D., Bramlett, R. K., Barnett, D. W., & Cox, F. (1994). Classification of risk status in kindergarten screening: A comparison of alternative measures. *Journal of Psychoeducational Assessment*, 12, 154-164.
- Kaminski, R. A., & Good, R. H. (1996). Towards a technology for assessing basic early literacy skills. *School Psychology Review*, 25, 215-227.
- Kelly, M. S., & Peverly, S. T. (1992). Identifying bright kindergartens at risk for learning difficulties: Predictive validity of a kindergarten screening tool. *Journal of School Psychology*, 30, 245-258.
- Laosa, L. M., & Siegel, I. E. (Eds.) (1982). *Families as learning environments for children*. New York: Plenum Press.
- Lichtenstein, R., & Ireton, H. (1991). Preschool screening for developmental and educational problems. In B. Bracken (Ed.), *The psychoeducational assessment of preschool children* (2nd. ed) (pp. 486-513). Boston, MA: Allyn and Bacon.
- Mantzicopoulos, P. Y. (1997). The relationship of family variables to Head Start children's preacademic competence. *Early Education and Development*, 8, 357-375.
- McLoughlin, C. S., & Rausch, E. (1990). Best practices in kindergarten screening. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-II* (pp. 455-467). Washington, D. C.: National Association of School Psychologists.
- Mouton, P. Y., & Tuma, J. M. (1988). Stress, locus of control and rated satisfaction in clinic and control mothers. *Journal of Clinical Child Psychology*, 17, 217-224.
- Patterson, G. R. (1988). Stress: A change agent for family process. In N. Garnezy & M. Rutter (Eds.), *Stress, coping, and development in children* (pp. 235-264). Baltimore, MA: The Johns Hopkins University Press.
- Pianta, R. C., Egeland, B., & Sroufe, L. A. (1990). Maternal stress and children's mental development: Prediction of school outcomes and identification of protective factors. In J. Rolf, A. Masten, D. Cicchetti, K. Neuchterlain, & S. Weintraub (Eds.), *Risk and protective factors in the development of*

PREDICTING FIRST GRADE ACHIEVEMENT

- psychopathology*. New York: Cambridge University Press.
- Reynolds, C. R., & Kamphaus, R. W. (1992). *Manual: Behavior Assessment System for Children*. Circle Pines, MN: American Guidance Service.
- Rutter, M. (1985). Family and school influences on behavioral development. *Journal of Child Psychology and Psychiatry*, 28, 349-368.
- Shepard, L. A. (1997). Children not ready to learn? The invalidity of school readiness testing. *Psychology in the Schools*, 34, 85 - 97.
- Siegel, I. E., McGillicuddy-DeLesi, A. V., & Goodnow, J. (Eds.) (1992). *Parental belief systems: The psychological consequences for children* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Teo, A., Carlson, E., Mathieu, P. J., Egeland, B., & Sroufe, L. A. (1996). A prospective longitudinal study of psychosocial predictors of achievement. *Journal of School Psychology*, 34, 285-306.
- Woodcock, R. W., & Johnson, M. B. (1990). *Woodcock-Johnson Tests of Achievement: Standard and Supplemental Batteries*. Allen, TX: DLM Teaching Resources.

Academic Achievement and Magnet Schools

Betty L. Dickson
Philander Smith College

Carolyn L. Pinchback
University of Central Arkansas

Robert L. Kennedy
University of Arkansas, Little Rock

Stanford Achievement Test, Eighth Edition, scores were collected in a large mid-southern urban district for elementary students who had attended the same school for three consecutive years. The sample comprised 560 students, equally divided among magnet and nonmagnet participants, as well as males and females, with an ethnic breakdown of two-thirds black and one-third white. An analysis of covariance and a Tukey-Kramer multiple comparison test revealed statistically significant differences among all four magnet/nonmagnet and black/white groups with adjusted means of 557 for the white magnet group; 572 for the white nonmagnet group; 584 for the black magnet group; and 595 for the black nonmagnet group. Eta squared effect sizes were small, however, ranging from 0.001 to 0.055.

The concept of magnet schools is not a new one. In fact Kirkpatrick (1994) reported that the magnet school concept originated with European grammar schools. Despite this possible heritage, magnet schools began developing only as recently as the early 1960s, as a result of social and political movements (Cooper, 1990). Conservatives saw magnets as allowing selectivity and specialness, while liberals viewed magnets as being able to bring black and white students together voluntarily (Cooper, 1990). Magnet schools were thus designed as a means of aiding large urban school districts in achieving desegregation by offering parents incentives to send their children to integrated schools. To reduce racial isolation in the public schools, magnet schools offered a curriculum based on a special theme, such as foreign languages, technology, health sciences, math, science, or performing arts programs; or special instructional approaches such as individualized education, open classrooms, or ungraded

schools (Adenika-Morrow, 1996; Bank & Spencer, 1997; Elliot, 1996; Flaxman, Guerrero, & Gretchen, 1997; Fowler & Corley, 1996). These themes are reflective of the forebears of specialty schools such as the Bronx School for Science, Boston Latin School, and Chicago's Lane Tech (Blank, 1984). According to Dickson (1989), large urban school districts where magnet schools evolved include Minneapolis (the first court-sanctioned one); Houston Independent; New York; Fort Worth; Los Angeles Unified; Flint, Michigan; Lowell, Massachusetts; and Milwaukee, Wisconsin.

Magnet schools even now continue to be one method of addressing racial isolation in schools. They have grown from being implemented in 14 school districts in 1976 to approximately 230 districts in 1994 (Steel & Levine, 1994). Although initially they could only be found in large urban school districts, now many small school districts have instituted magnet schools.

Henig (1995, p. 729) reported that "magnet schools more recently have been rhetorically linked to the goal of educational excellence." This statement suggests that magnets can have a positive effect on students' achievement, or as Black commented (1996, p. 35), they can provide "high-quality educational programs." Each of these programs is uniquely designed to attract students from across boundary lines, different economic classes, and different racial groups (Adenika-Morrow, 1996; Fullington, 1977; Woods, 1996). These incentives include curriculum innovation and high quality educational programs based on a special theme. Hence, these schools can offer the following benefits:

Betty L. Dickson is Chair, Department of Elementary Education, Philander Smith College, in Little Rock, AR. Carolyn L. Pinchback is a faculty member in the Department of Mathematics, University of Central Arkansas, in Conway, AR. Robert L. Kennedy is a faculty member in the Department of Educational Leadership, University of Arkansas, Little Rock, AR. An earlier draft of this paper was presented to the 1997 Annual Meeting of the American Educational Research Association in Chicago. Correspondence regarding this paper should be directed to Robert L. Kennedy, Department of Educational Leadership, University of Arkansas, Little Rock, 2801 S. University, Little Rock, AR 72204-1099 or by e-mail to rlkenney@ualr.edu.

- (1) more focused and advanced classes,
- (2) more funds and materials,
- (3) more qualified teachers with smaller classes,
- (4) fewer discipline problems, and
- (5) more whites in minority schools (Forte, 1998, p.9).

As an example of some of these magnet school benefits, Benjamin Banneker Computers Unlimited Elementary School in Kansas City, Missouri, found positive results with its emphasis on technology. Attendance improved and students worked at or above grade level (Richey, 1994). Another example occurred in Nigeria where magnet schools emphasizing giftedness offered facilities and resources so that students could "... realize their academic potential to the fullest" (Kolo, 1996, p. 79).

In 1980, the Achievement Goals Program in San Diego was devised to improve achievement throughout a group of schools. Part of its approach was moving away from an overemphasis on low-level skills. The project staff reported large gains in reading comprehension, language, and math at all grade levels (Levine & Ornstein, 1993). Dickson (1989) compared students' academic achievement in magnet schools versus that in nonmagnet schools. The Metropolitan Achievement Test, Sixth Edition (MAT6), scores of fifth grade magnet students in the Little Rock School District were compared to the MAT6 scores of fifth grade nonmagnet students in the district, over a three-year period. Significant differences were found in that students who attended the magnet schools scored higher in all areas on the MAT6 than students in nonmagnet schools. A few years later, Lewis (1990) studied Chicago Public Schools and concluded that "... on average, students of all racial groups in academic magnet schools and schools with competitive entry requirements, perform well above norms. ..." (p. i).

In a study similar to Dickson's, Moore (1991) interestingly found dissimilar results. Investigating three science and mathematics middle magnet schools over a four-year period, he reported that analysis of covariance results of Iowa Test of Basic Skills scores suggested "... a lack of convincing evidence of a math/science magnet program effect" (p. 23).

Barnes and Wesson (1994) examined the effects in Forrest City, Arkansas, of the discontinuation of federally funded magnet schools designed to eliminate racial imbalance and improve academic achievement in the elementary schools. The authors analyzed achievement differences in reading, language arts, mathematics, science, and social studies among minority and nonminority third and fourth grade students for two years. They found that white females scored somewhat higher than their counterparts. The study also revealed that federally

funded magnet schools had a significant effect on students' academic achievement.

In a 1995 report of a longitudinal study, Archbald said that magnet students generally had the highest performance when compared with neighborhood school students and predominately black transfer students. More recently, Gamoran (1996) found in a national survey that achievement was higher in public magnet schools than in public comprehensive schools. In fact, in science, reading, and social studies the differences were statistically significant. Gamoran noted that the study did not distinguish among different types of magnet schools nor did it investigate the possible effects of drawing students away from nonmagnet schools.

While these studies, for the most part, found students in magnet schools to have higher levels of academic achievement than their counterparts, there is still room for debate over how and whether magnet schools are effective for all racial groups (Archbald, 1995; Blank & Archbald, 1992; Hausman, Goldring, & Moirs, 1997; Steel & Levine, 1994; Thompson, 1995). Despite Black's (1996) suggestion that an ideal evaluation of magnet schools should include tracking achievement outcomes of students in magnet and nonmagnet schools, and despite the potential benefits and rapid growth of magnet schools in recent decades (NSBA, 1990), relatively few empirical studies have been available to compare student achievement in magnet and nonmagnet schools (Gamoran, 1996) or to support the claims of magnet programs (Hausman, Goldring, & Moirs, 1997). The purpose of this study was to compare the academic achievement of magnet and nonmagnet school students in both black and white elementary schools in a large mid-southern urban district.

Method

Participants

Data were collected by first identifying from school records all elementary students who were enrolled in the first grade of a large mid-southern school district during the 1993-94 school year. From this list, students who had attended the same school for three consecutive years (1993-94, 1994-95, and 1995-96) were selected. The sample comprised over 1,000 students, consisting of two-thirds black and one-third white, and evenly divided among male and female students. The magnet-elementary students (280 of the 1069 students in the sample) came from six magnet schools in the district whose themes included the arts, basic skills, high intensity learning, international studies/foreign languages, and math/science.

Membership in a magnet school in this district required the filing of an application and then selection on the basis of racial and school district participation

ACADEMIC ACHIEVEMENT AND MAGNET SCHOOLS

balance. That is, the racial mix in the magnets was comparable to the racial mix of the district as a whole. Also, there were three districts in the county which might send students to this district's magnet schools and a certain percentage of membership was reserved for each of these three districts. If the available slots were not taken by the other districts then they might be assigned to the remaining districts.

Because the magnet group was substantially smaller than the nonmagnet group, an equal-sized group of 280 students was randomly drawn from the nonmagnet group. The number of Hispanic and Asian students (13, or two percent, of the 560) was considered to be too small to be fairly representative of those groups, so they were excluded from the inferential analysis. Table 1 lists the demographic representation of the population and sample magnet and nonmagnet students in the district. In Table 2, a breakdown of the number of magnet students, by theme, is provided.

Instrument

The Stanford Achievement Test, Eighth Edition (SAT-8) scaled scores were collected from the spring (March/April) administrations of the first and third grades for each student for the total complete battery, which included language, listening, mathematics, reading, science, social science, spelling, and study skills. According to Brown (1992), the SAT-8 has a long history of use in the public schools and is one of the best norm-referenced

batteries currently available to assess achievement. Its purpose is to measure overall student achievement in the above areas.

According to the Stanford Achievement Test, 8th Edition, technical manuals (*Multilevel National Norms Booklet*, 1989, and *Technical Data Report*, 1990), the norm group consisted of 175,000 students in the spring 1988 administration and 135,000 from the following fall. The students were balanced within five percent of census percentages for high, medium, and low socioeconomic status; urban, suburban, and rural locales; and by ethnicity. Subscale reliability, as measured by KR-20, ranged from 0.75 to 0.94. Content validity was verified by a wide range of content experts comprising teachers, professors, and other recognized professionals. The content was matched to the half dozen most widely used texts in each topical area as well as to selected state curricular guidelines.

Scaled scores for the first grade ranged from the upper-200's to the mid-700's, while for the third grade, from the lower-300's to the lower-800's. Scaled score means were 538 for primary one and 599 for primary three. No standard deviations were provided for these scaled score means.

The means and standard deviations of the magnet/nonmagnet and black/white subgroups for the first and third grade scores are given in Table 3. The Pearson correlation between the first and third grade scores was 0.88.

Table 1
Comparison of Demographics of Population and Samples

	Population Magnets	Population Nonmagnets	Sample Magnets	Sample Nonmagnets
Race				
Black	177 (63%)	483 (61.2%)	177 (63%)	178 (64%)
White	96 (34%)	287 (36.4%)	96 (34%)	96 (34%)
Hispanic/Asian	7 (3%)	19 (2.4%)	7 (3%)	6 (2%)
Total	280 (100%)	789 (100%)	280 (100%)	280 (100%)
Gender				
	Population Magnets	Population Nonmagnets	Sample Magnets	Sample Nonmagnets
Male	147 (52.5%)	398 (50%)	147 (52.5%)	145 (52%)
Female	133 (47.5%)	391 (50%)	133 (47.5%)	135 (48%)
Total	280 (100%)	789 (100%)	280 (100%)	280 (100%)
Free or Reduced Lunch				
	Population Magnets	Population Nonmagnets	Sample Magnets	Sample Nonmagnets
Free/Reduced	113 (40%)	354 (45%)	113 (40%)	132 (47%)
Non-free/reduced	167 (60%)	435 (55%)	167 (60%)	148 (53%)
Total	280 (100%)	789 (100%)	280 (100%)	280 (100%)

Table 2
Magnet School Membership by Theme

School	N	%	Theme
1	47	16.8	Arts
2	54	19.3	Basic skills/Math-Science
3	33	11.8	International Studies/Foreign Languages
4	41	14.6	High Intensity Learning
5	56	20.0	Basic skills/Math-Science
6	49	17.5	Basic skills
Total	280	100	

Table 3
Subgroup Means and Standard Deviations
for SAT-8 Complete Battery

	Magnet			Nonmagnet		
	N	Mean	SD	N	Mean	SD
First Grade						
Black	174	531	29.6	174	523	29.5
White	96	563	34.1	96	550	38.1
Third Grade						
Black	177	572	29.3	174	565	27.5
White	96	605	34.0	95	599	38.0

Design

To compare the achievement levels of the magnet/nonmagnet and black/white groups, an analysis of covariance (ANCOVA) was chosen to account for possible initial differences in the first grade SAT-8 complete battery scores. The ANCOVA is a statistical method for examining data while using information about one or more independent variables to remove systematic differences caused by these covariates (Shavelson, 1988, p. 530). The main advantages of the ANCOVA over an analysis of variance (ANOVA) are that the ANCOVA can offer greater power and reduce the bias that can occur because of differences that exist between groups before a treatment is administered (Huitema, 1980, p. 13), or in this case, before the students were enrolled in the magnet program.

The dependent variable was the third grade scores; the factor variables were magnet/nonmagnet and race; and the covariates were first grade scores, gender, and whether the student participated in a free or reduced lunch program. The initial run revealed no statistically significant ($\alpha = 0.05$) effect due to gender. Since the cell sizes for the white students for the free or reduced lunch program were small (three to five students each), the cells were collapsed across gender to increase the remaining cell sizes. The covariates in the new model were then the first grade scores and free or reduced lunch. Unfortunately, the assumption of independence of the covariate and treatment, which in this case was magnet school

membership, could not be met since membership was self-selected which could affect the first grade scores. In fact, a comparison of the means for the first grade magnet and nonmagnet students in Table 3 does show a difference favoring the magnet students. It became necessary to consider an alternative design. Since the samples could be interpreted as having come from two populations in the sense that one population consisted of students interested in the magnet program and the other consisted of those students not interested in the program, a nonequivalent-group study design seemed appropriate. Kenny (1979) suggested an approach based on true pretest scores and background variables that removes the variability associated with the background variables (Huitema, 1980, p. 320). Basically, there are five steps in this method:

1. Regress the pretest on the group membership (magnet/nonmagnet and race) dummy variables and the background (free or reduced lunch) variable.
2. Calculate the predicted pretest scores from the model developed in step 1.
3. Estimate the reliability of the difference between the observed and predicted pretest scores by dividing the pooled within-group sum of the deviation cross-products for the observed and predicted pretest scores by the square root of the product of the pooled within-group sums of squares of the two sets of scores.
4. Estimate the background-corrected pretest true scores by adding the predicted score to the product of the reliability coefficient from step 3 times the difference of the observed and predicted scores, for each observation.
5. Use the set of scores estimated in step 4 and the background variables as covariates and the posttest (third grade scores) as the dependent variable in the usual ANCOVA.

The assumptions of normality and homoscedasticity were verified visually with box plots, probability density traces, histograms, and normal probability plots, and by using the D'Agostino Omnibus Normality of Residuals and the Modified-Levene Equal-Variance tests. No serious violations of these assumptions were observed. Linearity and homogeneity of regression slopes were investigated with a series of scatterplot overlays constructed with the true-score-adjusted pretest scores and the posttest scores for the magnet/nonmagnet, race, and free or reduced lunch program factors. Evidence of near-parallel linear trend lines was observed. A homogeneity of regression F test (Huitema, 1980, p. 44-47) was conducted to verify the observation. The test consists of computations of four sums of squares: within-groups, within-groups residual, individual residual, and heterogeneity of slopes; and calculation of the F -ratio

ACADEMIC ACHIEVEMENT AND MAGNET SCHOOLS

from the mean squares from the latter two sums of squares. The calculated *F*-statistic was not significant, indicating that the slopes of the regression lines were similar.

Results

The results of the ANCOVA are listed in Table 4. Eta squared, the ratio of the explained variance to total variance (Marascuilo & Serlin, 1988, p. 115), is included to provide an estimate of the effect size for each independent variable. The effect size, ranging from 0 to 1, is an indication of the degree of departure from what was hypothesized (Cohen, 1988, p. 10), which in this case was no effect.

Table 4
ANCOVA Results of First Grade SAT-8 Scores and Free or Reduced Lunch on Third Grade SAT-8 Scores by Race and Magnet School Membership

Source	DF	Sum of Squares	Mean Square	F	Prob Level	Eta Squared
F/R Lunch (Cov.)	1	5711	5711	21.6	<.01	0.009
1st Scores (Cov.)	1	320804	320804	1213.5	<.01	0.483
NonMag/Mag (A)	1	17341	17341	65.6	<.01	0.026
Race (B)	1	36674	36674	138.7	<.01	0.055
A x B	1	510	510	1.9	0.17	0.001
Error	529	139854	264			
Total (adj.)	534	664564				
Total	535					

Statistically significant main effects differences were observed between the magnet and nonmagnet [$F(.05, 1, 529) = 65.59$, $p = 0.000004$] as well as the black and white groups [$F(.05, 1, 529) = 138.72$, $p < 0.000005$]. Since there were only two groups in each case, the values of the means revealed the higher groups. The adjusted mean for the nonmagnet group was 583 and for the magnet group, 571. The adjusted mean for the black group was 590 and for the white group, 565. To determine differences among the magnet/white, nonmagnet/white, magnet/black, and nonmagnet/black groups, it was necessary to conduct a multiple comparisons test to determine where they might have occurred. According to the Tukey-Kramer test, useful when the groups are unequal in size, all four of the groups were statistically significantly different from each other. In order, the adjusted means were 557 for the white, magnet group; 572 for the white, nonmagnet group; 584 for the black, magnet group; and 595 for the black, nonmagnet group. Table 5 illustrates these differences.

Table 5
Tukey-Kramer Multiple-Comparison Test Comparing Magnet/Nonmagnet and Black/White Groups

Group	N	Mean	Different From Groups
Magnet/White (MW)	96	557	NMW, MB, NMB
NonMagnet/White (NMW)	95	572	MW, MB, NMB
Magnet/Black (MB)	174	584	MW, NMW, NMB
NonMagnet/Black (NMB)	170	595	MW, NMW, MB

Discussion

The findings seem to contradict those noted in the literature review, that is, that magnet schools usually have been found to have a positive relationship with achievement. For the groups in this study, the adjusted posttest scores for the students in the magnet schools were lower than those of the nonmagnets for both black and white groups. Even more surprising was the finding that the adjusted posttest scores of the black students were higher than those of the white students for both magnet and nonmagnet students. These findings are certainly contrary to most of the results reported in the literature. It is important to recognize, though, that the effect sizes were generally small and that there are inherent weaknesses in nonequivalent control group studies, with respect to both internal and external validity. Generalization may not be possible under these circumstances. Further studies are needed to verify or challenge these findings.

Although the results suggest that there was no significant impact of the magnet schools on performance--in fact, they did not fare as well--there may be some interest as to why there is a difference initially for the magnet students. Perhaps the magnet schools are more attractive to higher-achieving students, in which case they may be offering instruction or other academic incentives which could be transferred or replicated in nonmagnet schools and benefit the nonmagnet students. Archbald (1996) perhaps addressed the point when he found evidence indicating a statistically significant relationship between choice of magnet school and parents' educational level as well as proximity to the magnet schools. Hausman and Goldring (1997) found similar results with parents in their survey choosing magnet schools for academic reasons as well as for proximity of their homes to the schools.

The lack of substantive effects related to the magnet schools suggests that other concerns might better serve the interests of magnet school researchers. For example, are there aesthetic benefits to attending magnet schools? Is the school climate, for instance, more conducive to a supportive educational learning environment? Is there higher retention? Are there fewer disciplinary actions?

Are there benefits to the parents, teachers, or staff in the magnet schools? Are the benefits cost effective compared to expenditures in nonmagnet schools? These and other issues might well be addressed, perhaps in place of continuing to probe the perceived academic benefits of magnet schools.

References

- Adenika-Morrow, T. J. (1996, May). A lifeline to science careers for African-American females. *Educational Leadership*, 53(8), 80-83.
- Archbald, D. (1995, Spring). A longitudinal cohort analysis of achievement among elementary-magnet students, neighborhood-school students, and transfer students. *Journal of Research and Development in Education*, 28(3), 161-68.
- Archbald, D. (1996, Spring). SES and demographic predictors of magnet school enrollment. *Journal of Research and Development in Education*, 29(3), 152-162.
- Bank, B. J., & Spencer, D. A. (1997, March). *Effects of magnet programs on educational achievement and aspirations*. Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL). (ERIC Document Reproduction Service No. ED 412 614)
- Barnes, A., & Wesson, L. H. (1994, August). *Contrasting the Forrest City School District's magnet program operating with federal funding and without federal funding*. (ERIC Document Reproduction Service No. ED 380 888)
- Black, S. (1996). The pull of magnets. *The American School Board Journal*, 183(9), 34-36.
- Blank, R. K. (1984). The effects of magnet schools on the quality of education in urban school districts. *Phi Delta Kappan*, 66(4), 270-72.
- Blank, R. K., & Archbald, D. A. (1992, November-December). Magnet schools and issues of education quality. *Clearing House*, 66(2), 81-86. (ERIC Document Reproduction Service No. EJ 465 224)
- Brown, F. G. (1992). Review of the Stanford Achievement Test, eighth edition. In *The Eleventh Mental Measurements Yearbook*. J. J. Kramer & J. C. Conoley (Eds.). Lincoln, NE: Buros Institute of Mental Measurements, 861-863.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Cooper, B. S. (1990). An international comparison: Political forces for choice. In *Education reform: Making sense of it all*. S. B. Bacharach (Ed.). Boston: Allyn and Bacon.
- Dickson, B. L. (1989). *Students' Academic Achievement in a Magnet School versus Students' Academic Achievement in Non-Magnet Schools*. Unpublished dissertation, University of Arkansas, Fayetteville.
- Elliot, I. (1996). At home in the global village. *Teaching PreK-8*, 26(5), 34-38.
- Flaxman, E., Guerrero, A., & Gretchen, D. (1997). *Career development effects of career magnets versus comprehensive schools*. Washington, DC: Office of Vocational and Adult Education. (ERIC Document Reproduction Service No. ED 414 471)
- Forté, B. (1998, November). How magnet schools track students. *Mathematics Education Dialogues* 9. Reston, Virginia: National Council of Teachers of Mathematics.
- Fowler, R. C., & Corley, K. K. (1996, April). Linking families, building community. *Educational Leadership*, 53(7), 24-26.
- Fullington, G. G. (1977). Magnet schools: What are they? Where are they going? *OSSC Bulletin*, 21(3). (ERIC Document Reproduction Service No. ED 145 516)
- Gamoran, A. (1996, October). Do magnet schools boost achievement? *Educational Leadership*, 54(2), 42-46.
- Hausman, C. S., & Goldring, E. B. (1997, October/November). *Parent Involvement and Satisfaction in Magnet Schools: Do Reasons for Choice Matter?* Paper presented at the Annual Convention of the University Council for Educational Administration (Orlando, FL). (ERIC Document Reproduction Service No. ED 414 619)
- Hausman, C. S., Goldring, E. B., & Moirs, K. A. (1997, March). *Organizational capacity for school improvement: Teacher reports in magnet and nonmagnet schools*. Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL). (ERIC Document Reproduction Service No. ED 407 709)
- Henig, J. R. (1995, Summer). Race and choice in Montgomery County, Maryland, Magnet Schools. *Teachers College Record*, 96(4), 729-34.
- Huitema, B. E. (1980). *The Analysis of Covariance and Alternatives*. NY: Wiley-Interscience.
- Kenny, D. A. (1979). *Correlation and Causality*. NY: Wiley-Interscience. In Huitema, B. E. (1980). *The Analysis of Covariance and Alternatives*. NY: Wiley-Interscience.
- Kirkpatrick, D. W. (1994). Magnet schools/charter schools. In *Educational Innovation: An agenda to frame the future*. (Ed.). C. E. Greenawalt II. Lanham, Maryland: University Press of America, Inc.
- Kolo, I. A. (1996, December). Reflections on the development of gifted education in Nigeria. *Roeper Review*, 19(2), 79-81.

ACADEMIC ACHIEVEMENT AND MAGNET SCHOOLS

- Lewis, J. H. (1990, October). *Chicago's two public school systems: Standardized test results compared by racial/ethnic groups*. (ERIC Document Reproduction Service No. ED 404 419)
- Marascuilo, L. A., & Serlin, R. C. (1988). *Statistical Methods for the Social and Behavioral Sciences*. New York: W. H. Freeman and Company.
- Metz, M. H. (1986). *Different by design: The context and character of three magnet schools*. New York: Routledge & Kegan Paul.
- Moore, W. P. (1991, October). *Achievement and enrollment: Evaluation of the science and mathematics middle magnet schools*. (ERIC Document Reproduction Service No. ED 349 173)
- NSBA. (1990). *A Survey of Public Education in the Nation's Urban School Districts*. Alexandria, VA: National School Boards Association.
- Richey, E. (1994). Urban success stories. *Educational Leadership*, 51(7), 55-57.
- Shavelson, R. J. (1988). *Statistical Reasoning for the Behavioral Sciences*. Boston: Allyn and Bacon, Inc.
- Stanford Achievement Test Series, 8th ed.: Multilevel National Norms Booklet*. (1989). New York: Harcourt, Brace, Jovanovich.
- Stanford Achievement Test Series, 8th ed.: The Technical Data Report*. (1990). New York: Harcourt, Brace, Jovanovich.
- Steel, L., & Levine, R. (1994). *Educational Innovations in Multiracial Contexts: The Growth of Magnet Schools in American Education*. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, California Department of Education, Washington, DC, Office of the Under Secretary.
- Thompson, J. (1995). Choice--Kansas City style: Addressing equity from multiple perspectives. *Equity & Excellence in Education*, 28(3), 20-25.
- Woods, C. J. (1996). Hope for homeless students. *Educational Leadership*, 54(3), 58-60.

The Effects of Chess Instruction on the Mathematics Achievement of Southern, Rural, Black Secondary Students

James P. Smith and Bob N. Cage
University of Louisiana-Monroe

The purpose of this study was to determine the effects of 120 hours of chess instruction on the mathematics achievement of southern, rural, black, secondary students. Instruments used were the mathematics section of the California Achievement Test (CAT) (Level 20), Witkins's Group Embedded Figures Test (GEFT), the Guilford-Zimmerman Spatial Visualization Test (SV), and Naglieri's Nonverbal Ability Test (NNAT). A 2 X 2 analysis of variance of the pre-tests found no significant differences in the scores of the treatment group (11 females, 9 males) and the control group (10 females, 10 males) by group or gender. However, a 2 X 2 ANCOVA of the post-test results found a statistically significant difference in the scores. The means of the group receiving chess instruction were significantly higher than the control group means on the CAT, $F(1, 39) = 4.14, p < .043$; GEFT, $F(1, 39) = 6.09, p < .019$; SV, $F(1, 39) = 14.13, p < .001$; and the NNAT, $F(1, 39) = 4.72, p < .048$. Statistically significant correlations were also found between all instruments on the pre-test scores.

Background

Over the past twenty years several studies have found correlations between chess expertise among children and spatial ability (Horgan & Morgan, 1989; Van Zyl, 1991). Other studies (Fennema & Sherman, 1977; Person & Ferguson, 1989) have also confirmed a statistically significant correlation between mathematics and spatial ability. Christiaen (1976) and Van Zyl independently demonstrated that chess instruction significantly improved mathematics achievement of elementary and middle school students.

Christiaen (1976) conducted his study with 40 ten-year old, white, male, 5th graders in Belgium. Half of the students received one hour of chess instruction per week for 15 months. At the end of that time he tested both groups for cognitive development by Piagetian standards and standardized intelligence tests, and recorded their school results. The chess group outperformed the non-chess group in all comparisons, but the only statistically

significant difference was in school results. Van Zyl followed the progress of 160 South African middle school students, 80 chess players and 80 non-chess players, for a 3-year period. Initially, there was no statistically significant difference in the IQ scores of the two groups. Compared to the pre-test results, the total IQ scores of the chess playing group were significantly higher than the control group during the post-test. These results, although not proving chess was the causal agent of the improved IQ scores, certainly deserved some attention.

Horgan and Morgan (1989) conducted a study with 113 nationally ranked junior players to determine if a relationship existed between spatial ability and chess expertise. Among Horgan and Morgan's hypotheses were: children who excel at chess will measure above average on spatial abilities as measured by Raven's Progressive Matrices Test, and for children who excel at chess, chess skill will be highly correlated with spatial abilities. Both hypotheses were confirmed.

DeGroot's (1978) earlier work found that spatial ability did not correlate well with the skill of adult chess players; however, he found a correlation between chess skill in adults and logical thinking ability. It should be noted, though, that the adults used in DeGroot's research were all chess masters or grandmasters and were 32 years of age or older. This suggests an important question: At what age or at what stage of cognitive development do chess players begin to rely more on their logical thinking ability than on spatial ability? This question has never been properly addressed as no serious quantitative studies have been conducted with high school-aged chess players (D. D. Horgan, personal communication, October 24, 1996).

James P. Smith received his Ed.D. from Louisiana Tech University in Curriculum and Instruction. He is currently Interim Director of the Center for Educational Research and Services at the University of Louisiana-Monroe. Bob N. Cage received his Ph.D. from Iowa State University in Educational Leadership and Research. He is currently Professor of Educational Leadership and Coordinator, Ed.D. Program at the University of Louisiana-Monroe. This research was supported, in part, by a grant from the Louisiana Education Quality Support Fund. Correspondence regarding this article should be directed to: James P. Smith, 112 Strauss Hall, University of Louisiana-Monroe, Monroe, LA 71201 or via e-mail to pawne4@hotmail.com.

Numerous studies have also found a statistically significant correlation between mathematics achievement and general intelligence to the cognitive style construct of field-dependence/independence (FD/I) (Casey & Nuttall, 1995; Cooperman, 1980; Kagan & Zahn, 1975; Mrolsa, Black, & Hardy, 1987; Pearson & Ferguson, 1989; Perney, 1976). FD/I has also been shown to have a statistically significant correlation to spatial ability (MacLeod, Jackson, & Palmer, 1986; Witkin & Goodenough, 1981).

Much of the achievement gap in mathematics between genders and races can be directly attributed to the manner in which individuals perceive their environment, their relationship to their environment, and their ability to mentally manipulate objects (Kagan & Zahn, 1975; Pearson & Ferguson, 1989; Voyer, 1996). It is therefore no surprise that individuals with the lowest mathematics achievement scores also demonstrate the greatest degree of field dependence (Forns-Santacana, Amador-Campos, & Roig-Lopez, 1993; Perney, 1976; Voyer).

Field-independent individuals have been described as having focused attention and being able to notice and remember details. They prefer to do things according to a sequential theme. They show a preference for working alone. They are also adept at seeing differences between apparently similar experiences (Schmeck, 1988). Rather than being linear or sequential, the field-dependent individual's organizational schemes involve more random or multiple accessibility components, allowing numerous and varied associations between experiences. His/her thinking is more intuitive than that of the analytic person, and emphasis on conscious control of thought is less evident. Field-dependent individuals like group work and have demonstrated a preference for academic subjects and occupations that are people-centered (Chinien & Boutin, 1993). (For a more detailed treatment of the use of field-dependence measures, see Melancon & Thompson, 1987.)

It may be possible that chess instruction can improve spatial ability, improve logical thinking ability, and increase attention to detail among individuals who have not been previously exposed to chess. In turn, these improved abilities may aid the individual in improving mathematics achievement. The question this study specifically addressed was: Would chess instruction improve the mathematics achievement of rural, southern, black high school students?

Purpose and Rationale

This study addressed four hypotheses specific to the use of chess instruction and its effects on mathematics achievement, field-dependence/independence, spatial

ability and nonverbal reasoning ability of the treatment group. The hypotheses were as follows:

Hypothesis 1:

The mean score of the treatment group members will be statistically significantly higher than the mean score of the control group members in mathematics achievement as measured by the mathematics section of the CAT (Level 20).

Hypothesis 2:

The mean score of the treatment group members will be statistically significantly higher than the mean score of the control group members on the FD/I scale as measured by the GEFT.

Hypothesis 3:

The mean score of the treatment group members will be statistically significantly higher than the mean score of control group members in spatial ability as measured by the Guilford-Zimmerman Spatial Visualization Test.

Hypothesis 4:

The mean score of the treatment group members will be statistically significantly higher than the mean score of the control group members in nonverbal reasoning ability as measured by the Naglieri Nonverbal Ability Test.

Method

The subjects in this study were 40 students, ages 16 and 17, in a southern, rural high school. The treatment group consisted of 20 students, 11 female and 9 male. Because enrollment in the chess class was open to all juniors and seniors in the high school, control of gender and academic achievement of the subjects was not possible.

The control group consisted of 20 juniors and seniors randomly selected from the general student population. Members of the control group were not enrolled in the chess class and selected another elective class, outside the control of the researcher. Elective class choices consisted of Band, Choir, Journalism, Human Ecology, Spanish, French, Physics, Advanced Math, and Calculus. Also, members of the treatment and control groups were enrolled in mathematics courses based on personal choices or academic sequencing, both of which were beyond the researcher's control. Sixteen members of the treatment group had completed or were enrolled in Algebra I, Geometry, and Algebra II. Four treatment group members were enrolled in Advanced Math or Calculus. All members of the control group had completed or were enrolled in Algebra I, Geometry, and

Algebra II; and eight were enrolled in Advanced Math or Calculus. All participants in the study were of African-American descent. African-Americans were selected as the subject of this study because historically, they have demonstrated the greatest degree of field-dependence of all ethnic groups (Pemey, 1976).

The course material was divided into 24 separate lessons. Lessons varied in length from one to three class sessions depending on the complexity of the lesson. The classroom lessons were supplemented with homework assignments from a workbook, class discussions, and chess competition within the class where students were able to apply newly learned tactics. The students received approximately 120 hours of instruction over a 5-month period.

One of the key components of chess instruction in this study was the use of visual imagery rehearsal as a visualization technique. Research has shown that the use of mental imagery appears to facilitate meaningful learning (Alesandrini, 1982). Although Couch and Moore (1992) found that field-independent learners benefitted more than field-dependent learners from the use of imagery rehearsal, this technique was still utilized. Imagery rehearsal was accomplished by having students solve and discuss chess tactics problems that had been programmed into a computer and projected on a large screen. The chess software selected for this purpose allowed the students to advance the position or take back moves as needed to solve the problem.

After permanent rosters for classes were generated in late August, consent forms were given to the subjects and their parents. The study began with administration of the four pre-tests after the consent forms were returned. The school counselor administered tests over a 4-day period. Standardized instructions for each test were followed precisely.

Regular classroom instruction began following the pre-tests. Textbooks for the class included *Chess Rules for Students* by Bain (1994), *Chess Tactics for Students* by Bain (1993), and *Essential Chess Endings Explained Move by Move* by Silman (1992). The lessons taught on a daily basis were taken from *Comprehensive Chess Course (Vols. 1 & 2)* by Pelts and Alburt (1992).

Supplementary materials were also provided from *Winning Chess Tactics* by Seirawan (1992), and *Chess* by Polgar (1994).

Instrumentation

The instruments used in this study were selected because of their high validity and the reported high reliability of test-retest scores. The mathematics section of the California Achievement Test (Level 20) was used as the pre- and post-test for mathematics achievement. The observed test-retest reliability of the CAT during this study was $r = 0.95$.

Spatial ability was measured with the Guilford-Zimmerman Test of Spatial Visualization (SV). This test measures the subject's capacity to manipulate images mentally. Scores generated by this instrument also have a history of high reliability, $r = 0.93$ (Guilford & Zimmerman, 1997). Reliability of this instrument during the present study was $r = 0.93$.

FD/I was measured using the Group Embedded Figures Test (Witkin, Oltman, Raskin, & Karp, 1971). The GEFT scores have high reliability when compared to the Embedded Figures Test (0.82 for males, 0.63 for females). Test re-test reliability of the GEFT during this study was high, 0.93 for females and 0.83 for males.

Nonverbal reasoning ability was measured using the Naglieri Nonverbal Ability Test (Naglieri, 1997). This test is often used to determine giftedness among subjects who use English as a second language. The reported test-retest reliability is high (0.88). During this study the scores on the NNAT had a reliability of $r = 0.93$.

Pre-test Data Analysis

Pre-test data from this study were analyzed using the SPSS-X (SPSS Inc., 1988) computer program (see Table 1). A 2 X 2 analysis of variance (ANOVA) was used to determine if any significant differences existed between the mean scores of the treatment group and the mean scores of the control group on the CAT, GEFT, SV, or NNAT by group or gender. No significant differences were found in the means of the pre-test scores for any of the instruments.

BEST COPY AVAILABLE

Table 1
Pre-test and Post-test Results

Test	Group	Mean	Adjusted Mean	Min	Max	Standard Deviation
CAT Pre-test	Treatment	55.40	--	21	98	20.10
	Control	58.55	--	27	100	21.29
CAT Post-test	Treatment	61.25	62.85	40	101	17.43
	Control	61.60	60.12	31	97	20.44
GEFT Pre-test	Treatment	7.40	--	1	18	5.58
	Control	7.25	--	1	18	4.84
GEFT Post-test	Treatment	9.65	9.63	0	18	5.34
	Control	7.90	8.07	2	18	4.84
NNAT Pre-test	Treatment	17.95	--	5	35	7.48
	Control	17.20	--	4	35	8.44
NNAT Post-test	Treatment	21.50	21.01	7	38	7.52
	Control	18.45	18.91	8	35	7.87
SV Pre-test	Treatment	7.51	--	-2.75	23.75	7.95
	Control	7.36	--	-2.50	28.75	8.09
SV Post-test	Treatment	13.61	13.12	3.75	35.25	8.57
	Control	8.70	8.73	-2.00	35.25	8.73

Results

A 2 X 2 analysis of covariance (ANCOVA), comparing post-test means by group (treatment and control) and gender was used to test hypotheses. Pre-test scores for each individual instrument were used as the covariate. Effect sizes for the ANCOVAs were calculated using Hedge's *g* and included the use of adjusted mean scores. Levene's Test of Equality of Error of Variances was used to determine the homogeneity of regression assumption for each ANCOVA.

H₀₁: The mean score of treatment group members will be statistically significantly higher than the mean score of control group members in mathematics achievement as measured by the mathematics of the CAT (Level 20).

The mean score of the treatment group was significantly higher than the mean score of the control group on the post-test of the CAT, $F(1, 39) = 4.14, p < 0.05$ (see Table 2). Adjusted means were 62.85 for the treatment group and 60.12 for the control group. This result supported hypothesis 1.

The mean score of the male participants was also significantly higher than the mean score of the female

participants on the post-test of the CAT, $F(1, 39) = 6.02, p < .021$. Male post-test mean scores on the CAT improved dramatically (13.9% for treatment males, 8.5% for control males) compared to female post-test mean scores (8.2% for treatment females, 1.7% for control females). Interaction by group and gender was not significant.

Table 2
2 X 2 ANCOVA of CAT Post-test Scores
by Group and Sex

Source	df	F	Effect Size (Hedge's <i>g</i>)
Group (G)	1	4.14*	.180
Sex (S)	1	6.02*	.154
G X S	1	.99	.028
Covariate			
CAT Pre-test	1	875.40***	.962
S = within group error	39	(14.73)	

Note. Values enclosed in parentheses represent mean square error.

Levene's error variance, sig = .692.

* $p < .05$. *** $p < .001$.

THE EFFECTS OF CHESS INSTRUCTION ON THE MATHEMATICS

Ho₂: The mean score of treatment group members will be statistically significantly higher than the mean score of the control group members on the FD/I scale as measured by the GEFT.

The mean score of the treatment group was significantly higher than the mean score of the control group on the post-test of the GEFT, $F(1, 39) = 6.02, p < .019$ (see Table 3). The adjusted means were 9.63 for the treatment group and 8.07 for the control group. This result supported hypothesis 2. The difference in mean scores by gender was not significant. Interaction by group and gender was also not significant.

Table 3
2 X 2 ANCOVA of GEFT Post-test Scores
by Group and Sex

Source	df	F	Effect Size (Hedge's g)
Group (G)	1	6.02*	.420
Sex (S)	1	.13	.070
G X S	1	.005	.000
Covariate			
GEFT Pre-test	1	194.99***	.848
S = within group error	39	(3.83)	

Note. Values enclosed in parentheses represent mean square error. Levene's error variance, sig = .984.

* $p < .05$. *** $p < .001$.

Ho₃: The mean score of treatment group members will be statistically significantly higher than the mean score of control group members in spatial ability as measured by the Guilford-Zimmerman Spatial Visualization Test.

The mean score of the treatment group was significantly higher than the mean score of the control group on the post-test of the Guilford-Zimmerman Spatial Visualization Test, $F(1, 39) = 14.13, p < .001$ (see Table 4). Adjusted means were 13.12 for the treatment group and 8.73 for the control group. The difference in mean scores by gender was not significant. Interaction by group and gender was also not significant.

Ho₄: The mean score of treatment group members will be statistically significantly higher than the mean score of control group members in nonverbal reasoning as measured by the Naglieri Nonverbal Ability Test.

The mean score of the treatment group was significantly higher than the mean score of the control group on the post-test of the NNAT, $F(1, 39) = 4.72, p < .048$ (see Table 5). Adjusted means were 21.01 for the treatment group and 18.91 for the control group. The difference in

mean scores by gender was not significant. Interaction by group and gender was also not significant.

Table 4
2 X 2 ANCOVA of SV Post-test Scores
by Group and Sex

Source	df	F	Effect Size (Hedge's g)
Group (G)	1	14.13***	.332
Sex (S)	1	3.47	.170
G X S	1	.110	.003
Covariate			
SV Pre-test	1	159.36***	.820
S = within group error	39	(13.85)	

Note. Values enclosed in parentheses represent mean square error. Levene's error variance, sig = .649.

*** $p < .001$.

Table 5
2 X 2 ANCOVA of NNAT Post-test Scores
by Group and Sex

Source	df	F	Effect Size (Hedge's g)
Group (G)	1	4.72*	.288
Sex (S)	1	2.53	.214
G X S	1	.78	.052
Covariate			
NNAT Pre-test	1	57.42***	.628
GEFT Pre-test	1	7.53**	.181
S = within group error	39	(7.24)	

Note. Values enclosed in parentheses represent mean square error. Levene's error variance, sig = .09.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Correlation Analysis of Pre-test Data

Correlation analysis of pre-test data from this study was consistent with previous research (see Table 6). Numerous studies (Pearson & Ferguson, 1989; Voyer, 1996) have reported a significant positive correlation between spatial ability and mathematics achievement. Several studies have also found correlations between FD/I and general intelligence (Cooperman, 1980; Satterly, 1979; Mrolsa et al., 1987). Griffin and Franklin (1996) conducted a study of undergraduate college students and concluded that measures of FD/I can be used to predict college performance.

Table 6
Pre-test Correlations ($N = 40$)

Correlation Coefficients	1	2	3	4
1. CAT	--	.427**	.628***	.561***
2. GEFT		--	.743***	.643***
3. NNAT			--	.760***
4. SV				--

** $p < .01$. *** $p < .001$. (2-tailed).

Discussion

ANCOVA results for the CAT showed that the treatment group scored significantly higher than the control group in mathematics achievement, $F(1, 39) = 4.14$, $p < .043$. This was consistent with the results of Christiaen (1976) and Van Zyl (1991) in their studies comparing the academic achievement of chess playing students versus non-chess playing students. The notable difference in this example was that subjects in this study were high school students while Christiaen and Van Zyl studied middle school students. Christiaen also measured achievement by comparing grades the students received in their classes rather than using several standardized instruments.

The ANCOVA of the GEFT post-test scores showed that the treatment group scored significantly higher than the control group, $F(1, 39) = 6.02$, $p < .019$. Although field-dependence/independence has been shown to be stable over time (Witkin, Goodenough, & Karp, 1967), it should be pointed out that the researchers in that study were measuring and recording levels of FD/I and not attempting to bring about change. The effects of interventions to alter an individual's level of FD/I could not be found in the literature.

One of the inconsistent findings of this study was the similarity of GEFT scores between males and females. Several studies (Goodenough & Karp, 1961; Perney, 1976; Witkin & Goodenough, 1977) have repeatedly demonstrated the tendency of males to score significantly higher than females on tests of field-dependence/independence. However, in this study there was no significant difference between genders. This anomaly may have been a characteristic of the subjects used in the study.

The ANCOVA results of the SV, $F(1, 39) = 14.13$, $p < .001$, and the NNAT, $F(1, 39) = 4.72$, $p < .048$, indicated that the mean scores of the treatment group were significantly higher than the mean scores of the control group on both of those measures. Several studies have shown that strong correlations exist between scores on the SV and mathematics achievement (Burnett, Lane,

& Dratt, 1979), and scores on the GEFT and mathematics achievement (Mrolsa et al., 1987).

Scores on the GEFT, NNAT, SV, and tests of mathematics achievement are extremely dependent upon the perceptual skills of the test taker. The high correlation between these measures suggests they share a common factor. This common factor may be closely related to the skill needed to perform mental rotation activities. It is also reasonable to suggest that any activity that brings about an improvement on one instrument will have a corresponding positive effect on the other.

Despite the fact that the treatment group showed significant improvement in mathematics achievement, it is not reasonable to make the claim that chess instruction could be responsible for improvements in such fundamental mathematical skills as multiplication, the ability to correctly solve algebra problems, or to perform geometric proofs. It is well within reason to suggest, as did Van Zyl (1991), that chess instruction brought about changes in the subjects that produced more patience, perseverance, concentration, and creativity. These qualities could then be applied to other academic areas, such as mathematics.

According to Van Zyl (1991), a chess player's ability to reason is further developed when assessing moves and variations. This involves intense concentration, anticipation, classification, transcending, adoption of priorities, and increased logical thinking ability.

This study also implied that altering a student's perceptual ability can have significant effects on that student's cognition. However, no attempt has been made to explain the mechanism by which this change in cognition occurs.

A chess player's perseverance often determines victory or defeat at the board. When asked in a written survey what they enjoyed most about playing chess, 16 of 20 students responded, "Winning." Increased patience, perseverance, concentration and creativity among the students receiving chess instruction cannot be ruled out as a contributing factor in their improved mathematics scores.

One important factor that needs to be emphasized is that only one of the treatment subjects knew how to play chess before the study began. This study presented a clear picture of what chess intervention can do, as opposed to studies of post hoc analysis comparing a group of chess playing students against a group of non-chess playing students.

Future Research

Further study is needed to more adequately define and measure the underlying mechanism linking mathematics achievement, field-dependence, spatial ability, and nonverbal reasoning. Future research should include

additional instruments which measure self-esteem, perseverance, and student motivation. If, as Van Zyl (1991) claims, perseverance, motivation, and creativity are increased among chess players, these studies should attempt to determine if changes in field-dependence, spatial ability, and nonverbal reasoning account for the change in mathematics achievement scores, or if the change should be attributed to affective changes which were not measured quantitatively in the present study.

References

- Alesandrini, K. L. (1982). Imagery-eliciting strategies and meaningful learning. *Journal of Mental Imagery*, 6, 125-141.
- Bain, J. (1993). *Chess tactics for students*. Corvallis, OR: Learning Plus.
- Bain, J. (1994). *Chess rules for students*. Corvallis, OR: Learning Plus.
- Burnett, S. A., Lane, D. M., & Dratt, L. M. (1979). Spatial visualization and sex differences in quantitative ability. *Intelligence*, 3, 345-354.
- Casey, M. B., & Nuttal, R. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology*, 31, 697-705.
- Chinien, C. A., & Boutin, F. (1993). Cognitive style FD/I: An important learner characteristic for educational technologists. *Journal of Educational Technology Systems*, 21, 303-311.
- Christiaen, J. (1976). *Chess and cognitive development: An experimental psychological study of youths at the end of the primary school period*. Unpublished doctoral dissertation, Gent National University, Gent, Belgium. (Translated by S. Epstein).
- Cooperman, E. W. (1980). Field differentiation and intelligence. *The Journal of Psychology*, 105, 29-33.
- Couch, R. A., & Moore, D. M. (1992). The effects of imagery rehearsal, strategy, and cognitive learning style on the learning of different levels of instructional objectives. *International Journal of Instructional Media*, 19(1), 53-63.
- DeGroot, A. D. (1978). *Thought and choice in chess*. The Hague: Mouton.
- Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization and affective factors. *American Educational Research Journal*, 14, 51-71.
- Forns-Santacana, M., Amador-Campos, J. A., & Roig-Lopez, F. (1993). Differences in field dependence-independence cognitive style as a function of socioeconomic status, sex, and cognitive competence. *Psychology in the Schools*, 30, 176-186.
- Goodenough, D. R., & Karp, S. A. (1961). Field dependence and intellectual functioning. *Journal of Abnormal and Social Psychology*, 63, 241-246.
- Griffin, R., & Franklin, G. (1996). Can college academic performance be predicted using a measure of cognitive style? *Journal of Educational Technology Systems*, 24, 375-379.
- Guilford, J. P., & Zimmerman, W. S. (1981). *Manual of instructions and interpretations*. Palo Alto, CA: Consulting Psychologists Press.
- Horgan, D. D., & Morgan, D. (1989, August). *Experience, spatial abilities, and chess skill*. Paper presented at the American Psychological Association meeting, Atlanta, GA. (ERIC Document Reproduction Service No. ED 305 145)
- Kagan, S., & Zahn, G. L. (1975). Field dependence and the school achievement gap between Anglo-American and Mexican-American children. *Journal of Educational Psychology*, 67, 643-650.
- MacLeod, C. M., & Jackson, R. A., & Palmer, J. (1986). On the relation between spatial ability and field dependence. *Intelligence*, 10, 141-151.
- Melancon, J. G., & Thompson, B. (1987, November). *Measuring characteristics of a test of field-independence: Literature review and development of the Finding Embedded Figures Test*. Paper presented at the annual meeting of the Mid-South Educational research Association, Mobile. ERIC Document Reproduction Service No. ED 292 823)
- Mroska, H., Black, W. L., & Hardy, C. A. (1987). Cognitive learning style and achievement in mathematics. *Journal of Instructional Psychology*, 14, 26-28.
- Naglieri, J. A. (1997). *Naglieri nonverbal ability test multilevel technical manual*. San Antonio, TX: The Psychological Corporation Harcourt Brace Jovanovich, Inc.
- Pearson, J. L., & Ferguson, L. R. (1989). Gender differences in patterns of spatial ability, environmental cognition, and math and English achievement in late adolescence. *Adolescence*, 24, 421-431.
- Pelts, R., & Albut, L. (1992). *Comprehensive chess course* (Vols. 1 & 2)(3rd ed.) New York: Chess Information and Research Center.
- Perney, V. H. (1976). Effects of race and sex on field dependence independence in children. *Perceptual and Motor Skills*, 42, 975-980.
- Polgar, L. (1994). *Chess: 5334 problems, combinations, and games*. New York: Black Dog & Leventhal Publishers, Inc.

- Satterly, D. J. (1979). Covariation of cognitive styles, intelligence, and achievement. *British Journal of Educational Psychology*, 49, 179-181.
- Schmeck, R. R. (1988). Strategies and styles of learning. In R. R. Schmeck (Ed.), *Learning strategies and learning styles* (pp. 317-347). New York: Plenum Press.
- Seirawan, Y. (1992). *Winning chess tactics*. Redmond, WA: Tempus Books of Microsoft Press.
- Silman, J. (1992). *Essential chess endings explained move by move*. Dallas: Chess Digest, Inc.
- SPSS Inc. (1988). *SPSS user's guide* (3rd ed.). Chicago: SPSS Inc.
- Van Zyl, A. S. A. J. (1991, October). *Chess promotes the differentiation of latent intellectual potential*. Paper presented at the 10th annual congress of the Psychological Association of South Africa, Pretoria, South Africa.
- Voyer, D. (1996). The relation between mathematical achievement and gender differences in spatial abilities: A suppression effect. *Journal of Educational Psychology*, 88, 563-571.
- Witkin, H. A., & Goodenough, D. R. (1977). Field dependence and interpersonal behavior. *Psychological Bulletin*, 84, 661-689.
- Witkin, H. A., & Goodenough, D. R. (1981). *Cognitive styles: Essence and origins*. New York: International Universities Press.
- Witkin, H. A., Goodenough, D. R., & Karp, S. A. (1967). Stability of cognitive style from childhood to young adulthood. *Journal of Personality and Social Psychology*, 7, 291-300.
- Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). *Group embedded figures test manual*. Palo Alto, CA: Consulting Psychologists Press, Inc.

Changing Teacher Education Students' Attitudes toward Using Corporal Punishment in the Classroom

Marlynn M. Griffin
Georgia Southern University

Daniel H. Robinson
University of Louisville

Holli M. Carpenter
Georgia Southern University

There is overwhelming evidence that corporal punishment does not produce long-term changes in behavior and, worse still, may cause physical and psychological damage to the recipient of the punishment (Cryan, 1987; Larzelere, 1986; Straus & Donnelly, 1993; Straus & Kantor, 1994; Turner & Finkelhor, 1996). In a 1992 survey, Abrahams, Casey, and Daro found that only 38% of teachers in the South found abolishing corporal punishment in the classroom to be of above average importance, as compared with 68% of teachers in other regions. In two Southern universities, teacher education students wrote a literature review with empirical references on the pros and cons of corporal punishment. Using pretest and posttest attitude measures, this study found a decrease in students' support for corporal punishment.

Despite an ever-increasing number of classroom management alternatives and programs, some educators and prospective educators still advocate the use of corporal punishment as an effective means for decreasing inappropriate behavior. Punishment involves the presentation of an aversive stimulus in an attempt to decrease the immediately preceding behavior. Corporal punishment is often used synonymously with spanking or paddling (Block, 1994; Gursky, 1992; Hyman, 1993; Vockell, 1991), "but more broadly defined, it is any punishment that inflicts bodily pain for disapproved behavior" (Gursky, 1992, p. 18). Some school districts have even expanded this definition to include disciplinary techniques that some teachers may not consider to be harsh, such as making a child stand for an extended period of time, requiring a student to run in place for a lengthy time period, or depriving a child of recreation (Hyman, 1993; Ryan, 1994).

Because the purpose of punishment is to decrease behavior, for corporal punishment to be effective, it too

presumably must decrease behavior. Yet the data on this disciplinary technique show that this is often not the case. There is overwhelming evidence that corporal punishment does not produce long-lasting changes in behavior, and, worse still, may cause physical and psychological damage to the recipient of the punishment (Cryan, 1987; Larzelere, 1986; Straus & Donnelly, 1993; Straus & Kantor, 1994; Turner & Finkelhor, 1996). Perhaps it was for these reasons that the United States military banned the practice of corporal punishment back in 1874 (Heymann, 1991).

Some side effects of corporal punishment include running away or truancy (McCown, Driscoll, & Roop, 1996), fear of the teacher and/or school, high levels of anxiety (Biehler & Snowman, 1997), feelings of helplessness and humiliation, and aggression and destruction at home and at school (Cryan, 1995). Heymann (1991) found that 1,157,270 American children are corporally punished each year and of these, nearly 20,000 are seriously injured in the process.

In many instances the student who receives corporal punishment receives it repeatedly over time (Block, 1994). In the best-case scenario, the behavior for which a student is corporally punished is temporarily suppressed, but this suppression of misbehavior does not ensure that students will behave appropriately in the future or even that they will know what behavior is expected of them (Biehler & Snowman, 1997; Block, 1994). If any technique to decrease behavior is not accompanied by an explanation of appropriate or expected behavior, it is

Marlynn M. Griffin is an Associate Professor of Educational Foundations at Georgia Southern University. Daniel H. Robinson is an Assistant Professor of Educational Psychology at the University of Texas at Austin. Holli M. Carpenter is a recent graduate of Georgia Southern University. Correspondence relevant to this article should be addressed to Marlynn M. Griffin, Curriculum, Foundations and Research, P.O. Box 8144, Georgia Southern University, Statesboro, GA 30460-8144 or by e-mail to mgriffin@gasou.edu.

unreasonable to expect students to demonstrate acceptable behavior.

Despite the serious limitations to using corporal punishment, parents and teachers alike advocate its use in the classroom. Furthermore, only approximately half the states in the United States have prohibited its use in the classroom (Block, 1994; Hyman, 1993; Richardson, Wilcox, & Dunne, 1994), despite the fact that among developed countries (e.g., all of Europe, Japan, China), the U.S., Canada, and Australia alone still allow paddling. Additionally, many national organizations, such as the National PTA, the National Education Association, the American Medical Association, and the American Bar Association, have all spoken out against corporal punishment (Gursky, 1992). Yet, a 1988 poll (Gallup & Elam) found that 50% of the public approved of paddling in the schools, 45% opposed it, and 5% weren't sure how they felt. A survey published in *Parents Magazine* (Groller, 1989) indicated that 38% of parents nationwide approved of paddling to discipline school children, whereas in Southern regions 53% of parents approved of this disciplinary practice. In a 1992 survey of teachers, Abrahams, Casey, and Daro found that only 38% of teachers in the Southern United States found abolishing corporal punishment in the classroom to be of above average importance, as compared with 68% of teachers in other regions of the U.S.

As indicated above, approval ratings of corporal discipline are disproportionate across regions of the United States, with Southern states indicating a greater acceptance of this practice. The Department of Education indicated that the top ten states in number of reported paddlings were all Southern states: Mississippi, Arkansas, Alabama, Tennessee, Texas, Georgia, Louisiana, Oklahoma, South Carolina, and Kentucky (Waters, 1993). Heymann (1991) found similar results with the exception that Florida was listed among the top ten states instead of Kentucky.

The first two authors of this paper have taught introductory educational psychology courses for several years in Southern universities to teacher education students. In open discussion about classroom management, the authors have found that many students espouse corporal punishment as an appropriate means for changing behavior mostly because "I got spanked and I turned out okay" or "Schools will fall to anarchy without the paddle." These reasons mirror those used by advocates of corporal punishment across the country (Block, 1994; Gursky, 1992). When presented with negative evidence concerning the effectiveness of corporal punishment by the instructors, many students become defensive and cling even more strongly to their belief that it is appropriate to paddle public school students.

The purpose of this study was to investigate a method for changing teacher education students' attitudes about the appropriateness of using corporal punishment in the classroom. Although numerous studies were found that showed negative effects of corporal punishment, none to the authors' knowledge have attempted to change the attitudes of persons who support it. Specifically, we thought that if students could become informed about the negative consequences of corporal punishment through discovery rather than listening to instructors preach about it, they might become less defensive and more likely to change their attitudes.

Method

Participants and Design

Ninety-two students enrolled in two undergraduate educational psychology courses at two different Southern state universities (Georgia and Mississippi) participated for course credit. The classes consisted of approximately 70% female, 30% male, and about 15% African-American students. Students were randomly assigned in equal numbers to either a treatment or a control group. From the Georgia university, there were 30 students in the treatment group and 26 students in the control group, whereas the Mississippi university had 29 students in each group. To ensure that every student was exposed to a potentially beneficial treatment, a crossover design was used where the treatment and control groups reversed roles after the first posttest.

Materials

Students in the treatment group at both universities were required to write a literature review on "the pros and cons of corporal punishment in America's schools." Each student was required to include at least five references to empirical research articles in the paper. Students were encouraged to use computer databases such as ERIC and PSYCHLIT, and in the last paragraph of the paper, they were required to summarize the findings by stating whether or not corporal punishment "works." The papers were to be about two pages in length, typed, double-spaced, and students were instructed not to work with notes to discuss their assignment with their classmates.

Students in the control group were required to write a different paper. At the Georgia university, control students wrote on the topic of "what constitutes academic dishonesty," whereas at the Mississippi university, they wrote on the topic of "good teachers I have known." Neither assignment required references, and they too were to be about two pages in length.

The questionnaire consisted of five items. Each item consisted of a statement and five options that best

characterized the student's position (strongly agree, agree, undecided, disagree, strongly disagree). One of the items read, "As a teacher, I am in favor of using corporal punishment as a disciplinary method." The other four items that did not refer to corporal punishment were listed simply to help mask the purpose of the study.

The following reliability and validity evidence was obtained for the one-item corporal punishment attitude measure. Control group students' responses to the item before and after the assignment were correlated to provide a measure of test-retest reliability. The Pearson correlation was .794. Face validity for the item was judged to be sufficient by the authors.

Procedure

On the first day of class, students were given the questionnaire and completed it in about five minutes. Only students' responses to the corporal punishment item were recorded. Students were randomly assigned to be in either the treatment or control group.

Approximately one month later, students were given their respective assignments. Students in the treatment group received the corporal punishment assignment, whereas students in the control group received the alternative assignment (either "good teachers" or "academic dishonesty"). They were given one week to complete the assignment. After one week, students turned in their assignments and completed the questionnaire for a second time.

Students then received a second assignment. This time students in the treatment group received the alternative assignment and students in the control group received the corporal punishment assignment. After one week, students turned in their second assignments, which was followed by a third completion of the questionnaire.

Results

Students' responses were scored as "strongly agree" = 5, "agree" = 4, "undecided" = 3, "disagree" = 2, and "strongly disagree" = 1. A mixed ANOVA was conducted on the corporal punishment scores with experimental group as the between-subjects factor and testing occasion as the within-subjects factor. Table 1 presents means and standard deviations for the scores. The main effect of group was not significant, $F(1, 90) = .66$, $MSE = 3.42$, $p = .42$. The main effect of testing, however, was significant, $F(2, 180) = 4.25$, $MSE = .62$, $p < .05$, $\eta^2 = .045$. A Tukey Honestly Significant Difference test was used to follow up this effect. For all students, their attitudes toward corporal punishment at the beginning of class ($M = 3.01$, $SD = 1.34$) were more favorable than they were after completing the assignments ($M =$

2.67, $SD = 1.20$). The interaction effect of group by testing was not significant, $F(2, 180) = .99$, $p = .37$.

Table 1
Means (and Standard Deviations) for Corporal Punishment
Acceptability Ratings (1 = Strongly Disagree, 5 = Strongly Agree)

Time that Teacher Questionnaire was Administered	Group		Control	
	Treatment ($n = 46$)		($n = 46$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Beginning of term	3.00	(1.38)	3.02	(1.31)
After completing first assignment	2.65	(1.25)	3.00	(1.12)
After completing second assignment	2.59	(1.15)	2.76	(1.25)

To determine whether the groups of students from Southern universities were different in terms of their initial attitudes towards corporal punishment than students from other regions of the U.S., the same questionnaire was administered to 24 undergraduates enrolled in an educational psychology course at the University of South Dakota. The mean score on the corporal punishment item was 2.50 as compared to 3.01 for the Southern students, providing support for the earlier claim that corporal punishment is more accepted in Southern states.

After all papers were completed, the instructors encouraged class discussion about the issues mentioned in the students' papers. Students were eager to discuss the results of their research, particularly as they pertained to the findings on corporal punishment. Students reported that they were surprised at the negative consequences of corporal punishment and were also surprised at how easy it was for teachers to punish students too severely when inflicting corporal punishment. Several also indicated that they were less likely to utilize corporal punishment after completing the assignment than they were before reading the research.

Discussion

The results of this experiment indicate that students' support for corporal punishment decreased over the course of the semester. Unfortunately, we were unable to obtain a statistically significant experimental effect to demonstrate that this change in attitude was due solely to the literature review paper the students completed. This effect failed to reach statistical significance due to low power, attributable to both a small effect size ($\eta^2 = .045$) and sample size ($n = 46$ in each group). Once all 92 students were considered in the overall effect of testing occasion, the change in their attitudes from the beginning of the semester to the end was statistically significant.

Because this research was designed to simply decrease Southern students' support for using corporal

punishment in the classroom, we were less concerned with exactly what caused the attitude change and more concerned with the fact that it did change over the course of the semester. The seemingly minuscule effect (a change from 3.01 to 2.67 on a five-point attitude scale) may result in major consequential effects if these pre-service teachers are influential in eventually abolishing the practice of corporal punishment.

The authors encourage teacher education instructors to consider using this intervention if they are interested in changing students' attitudes towards corporal punishment. It can be easily achieved as an assignment for the entire class. We also recommend discussion about corporal punishment *after* the assignment, rather than before. As previously mentioned, our experiences had been very negative when attempting to discuss corporal punishment without students having first explored the scientific literature. We found students to be very emotional about their beliefs and resistant to change. Once students discover the overwhelming scientific evidence regarding the negative effects of corporal punishment, they will be less likely to believe all arguments concerning corporal punishment must be based solely on personal experiences.

References

- Abrahams, N., Casey, K., & Daro, D. (1992). Teachers' knowledge, attitudes, and beliefs about child abuse and its prevention. *Child Abuse & Neglect*, 16, 229-238.
- Biehler, R. F. & Snowman, J. (1997). *Psychology applied to teaching* (8th ed.). Boston: Houghton Mifflin Company.
- Block, N. (1994). Paddling strikes out. *The American School Board Journal*, 181(9), 40-41.
- Cryan, J. R. (1987). The banning of corporal punishment in child care, school and other educative settings in the United States. *Childhood Education*, 63(3), 146-153.
- Cryan, J. R. (1995, Spring). The banning of corporal punishment. *Dimensions of Early Childhood*, 36-37.
- Gallup, A., & Elam, S. (1988, September). The 20th annual Gallup poll on the public's attitudes toward the public schools. *Phi Delta Kappan*, 70, 33-46.
- Groller, I. (1989, March). Spanking in the schools? *Parents*, 64, 39.
- Gursky, D. (1992, February). Spare the child? *Teacher Magazine*, 3(5), 17-19.
- Heymann, T. (1991). *The unofficial U.S. census*. New York: Fawcett Columbine.
- Hyman, R. T. (1993, November). *Corporal punishment: Just what is it and what should we do about it?* Paper presented at the annual meeting of the National Organization on Legal Problems of Education, Philadelphia, PA.
- Larzelere, R. E. (1986). Moderate spanking: Model or deterrent of children's aggressive behavior in the family? *Journal of Family Violence*, 1, 27-36.
- McCown, R., Driscoll, M., & Roop, P. (1996). *Educational psychology: A learning-centered approach to classroom practice* (2nd ed.). Needham Heights, MA: Allyn & Bacon.
- Richardson, R. C., Wilcox, D. J., & Dunne, J. (1994). Corporal punishment in schools: Initial progress in the Bible belt. *Journal of Humanistic Education and Development*, 32, 173-182.
- Ryan, F. J. (1994, Winter). From rod to reason: Historical perspectives on corporal punishment in the public school, 1642-1994. *Educational Horizons*, 72(2), 70-77.
- Straus, M. A., & Donnelly, D. A. (1993). Corporal punishment of adolescents by American parents. *Youth and Society*, 24, 419-442.
- Straus, M. A., & Kantor, G. K. (1994). Corporal punishment of adolescents by parents: A risk factor in the epidemiology of depression, suicide, alcohol abuse, child abuse, and wife beating. *Adolescence*, 29, 543-561.
- Turner, H. A. & Finkelhor, D. (1996). Corporal punishment as a stressor among youth. *Journal of Marriage and the Family*, 58, 155-166.
- Vockell, E. (1991). Corporal punishment: Pros and cons. *Clearing House*, 64, 278-283.
- Waters, R. (1993, October 12). "Mom, the teacher hit me:" The dirty little secret in school. *Family Circle*, 106(14), 150-151, 192-196.

A Comparison of Attitudes and Expectations of Teachers and Peers Towards Rural Gifted Female Adolescents

Deborah Bretz
George Jenkins High School

Neelam Kher and Lorna J. Lacina-Gifford
Northwestern State University (LA)

Gifted females face the complex burden of dealing with conflicting expectations in society. The purpose of this study was to compare the expectations of teachers and peers for gifted female adolescents and to discuss the implications of these expectations. Survey data were obtained from 28 teachers, 11 male and 17 female, of both gifted and non-gifted students in a rural school setting and from 34 peers, 22 male and 12 female, in the same rural setting. The peer sample consisted of seventh- and eighth-grade students found in two heterogeneously grouped classrooms. Significant differences existed between the expectations of teachers and peers for gifted female adolescents on 4 of 17 dimensions addressed. This study supports the suggestion that gifted females face a dilemma of conflicting societal expectations. It is recommended that special guidance be provided to help gifted female adolescents deal with the effects of these conflicting expectations.

Introduction

During the last few decades there have been changes in the roles and expectations that society accepts and reinforces for females. Unfortunately, the literature related to gifted females reflects that these changes may not be as extensive or far reaching as one might expect (Johnson & Lewman, 1990; Kravetz, 1984; Leung, Conoley, & Scheel, 1994; Louis & Lewis, 1992; Luftig & Nichols, 1991). According to Hingham and Navarre (1984), the deck appears stacked against the gifted female adolescent. Young women often find themselves confused by societal messages encouraging them to place less emphasis on academic and career advancement in order to maintain traditional feminine roles of dependence on others and non-assertiveness (Clickenbeard, 1991; Gilligan, Lyons, & Hammer, 1990; Kleinsasser, 1986; Randall, 1997; Sisk, 1987). Despite this confusion, many researchers (Howard-Hamilton & Robinson, 1991; Kerr, 1985) have overlooked differences related to gender. In fact, research on the development of gifted females is sparse, thus providing incomplete conclusions since it centers mainly on studies of males (Kerr, 1985). Therefore, gifted

females may be especially at-risk in a world where standard setting development and research is often focused on males (Kline & Short, 1991).

Review of the Literature

It is at a young age that gifted females begin to receive the message that so far as risk-taking and their own safety are concerned, they should trust others before themselves. Risk-taking, self-trust and independence, however, are necessary for the development of superior intellect (Clark, 1992). In the home, the dilemma of conflicting expectations for gifted females begins at an early age. Although, parents want strong, self-sufficient children, they either intentionally or unintentionally inculcate their small daughters with the notions that they are somehow dependent rather than independent individuals (Johnson & Lewman, 1990).

As girls progress through school, many overt clues influence their role expectations. For example, preschool teachers reinforce the same type behavior noted to occur in the home by allowing little girls to stay close at hand and by speaking to them quietly (Sadker & Sadker, 1986). Boys, however, typically get more attention when they misbehave (American Association of University Women, 1991; Sadker, Sadker & Klein, 1991). Boys are also encouraged to explore toys and materials, whereas girls are shown the same material passively (Clark, 1992; Sadker & Sadker, 1994). Teacher behaviors that reinforce stereotyped expectations may be found throughout gifted females' school experiences (Howard-Hamilton & Robinson, 1991; Sadker & Sadker, 1986, 1994; Sadker et al., 1991).

Deborah Bretz, a recent masters graduate from Northwestern State University in Natchitoches, Louisiana, is a teacher at George Jenkins High School in Lakeland, Florida. Neelam Kher is Professor of Educational Psychology at Northwestern State University, and Lorna J. Lacina-Gifford is Professor of Education at Northwestern State University. Please address correspondence regarding this article to Lorna J. Lacina-Gifford, College of Education, Northwestern State University, Natchitoches, LA 71497 or by e-mail at giffordl@alpha.nsula.edu.

Several studies at the national level suggest that the middle grades can be a time of significant decline in the self-esteem and academic achievement for girls (AAUW, 1991; Backes, 1994; Orenstein, 1994). This is particularly true in the area of science achievement where the decline is much more evident for girls than for boys (Backes, 1994). Thus, parents and teachers need to make identifying the special needs of female adolescents a high priority at home and at school.

Gifted females are particularly vulnerable during the adolescent years. Society gives adolescent females mixed and confusing messages (Gilligan et al., 1990). Presently, gifted females are encouraged to contribute more of their abilities and talents to society. Many careers are available for their choosing and higher levels of advancement and success are opening for women. However, the societal belief in male superiority still exists in many professions. When girls observe that women hold positions of less status than men in society, it could lead girls to infer that their role is less important than that of boys (Debold, 1995).

Expectations of Society

Gilligan (1982) suggests that girls very early in life are guided through a different socialization process than boys. According to Erickson (1950) and Angrist (1969), by age four, boys begin to break patterns of dependence and move toward autonomy and independence. Girls, however, continue to be nurtured toward relational values such as intimacy and empathy. Thus, by early adolescence, girls already experience value conflicts in a society which values the independence typically encouraged in males. In fact, parents even attribute success of gifted females to hard work and luck, whereas they attribute success of males to special ability (Dickens & Cornell, 1993; Heller & Ziegler, 1996; Randall, 1997). Female role models, be they parental, educational or career oriented, are often seen in one of two ways. They are seen either as feminists balking at the traditional male orientation or as compromising themselves by playing according to male gender rules and trying to meet conflicting expectations in search of self-fulfillment (Kline & Short, 1991).

Junior high school is a crucial time for females, because attitudes begin to change with the emergence of negative feelings, low self-confidence and perfectionistic tendencies. This suggests the beginnings of emotional vulnerability (Kline & Short, 1991). Research shows that as gifted girls mature, they feel pressure to adopt certain goals, values and behaviors commensurate with society's preference for traditionally male independence and competitiveness (Horner, 1987). They also develop strong preferences and needs for relationship development. When goals and potentials conflict with the development

of relationships, gifted girls will choose to maintain positive relationships (Kline & Short, 1991).

Teacher Expectations

Gender affects teacher expectations and consequently, responses of gifted children in the classroom. Dweck and Bush (1976) found that teachers attributed poor performance of gifted females to a lack of ability, whereas poor performance on the part of gifted males was attributed to a lack of effort. Leroux (1986) found gifted adolescent females respond to learning experiences differently from their gifted male counterparts. Many parents still rely on gender to form expectations and perceptions about their children. Girls are often encouraged to engage in passive play activities, whereas boys are encouraged to create and build (Castenell, 1983; Howard-Hamilton & Robinson, 1991; Kerr, 1991; Randall, 1997).

Cooley, Chauvin, and Karnes (1984) compared the attitudes of male and female teachers towards gifted females. The survey of teachers highlighted the importance of teachers' attitudes towards sex role behaviors in the activities available to each sex in the classroom. Both male and female teachers saw gifted boys as more competent than gifted girls in problem solving and in critical and logical thinking skills. They saw gifted girls as more competent in creative writing. Cooley et al. (1984) concluded that perceptions regarding females are changing. However, the male teachers surveyed still had a tendency to see female students' characteristics more traditionally.

Peer Expectations

One of the most independent aspects of a child's life and development is that of the relationships she establishes with peers. Children's relationships and interactions with peers, whether positive or negative, are extremely important and influential. As children grow into adolescents, they interact increasingly independent of adults and form unique standards and expectations of behavior (Luftig & Nichols, 1991). Adolescence represents a time of change and adjustment in a young person's life.

The perception of gifted children's behaviors and the assigning of equitable social status by non-gifted peers becomes important in the social realm of schools. Luftig and Nichols (1991) examined the social status of gifted students as rated by their non-gifted peers. Gifted boys were found to be the most popular ability and gender group. Gifted girls were found not to be as socially accepted as their non-gifted peers, but neither gifted group was rejected by their peers or non-gifted gender counterparts (Luftig & Nichols, 1991; Randall, 1997). The conflicting societal and school messages highlighted in this study may be putting gifted girls at-risk in terms of academic achievement and psychological well being.

Purpose

Gifted females face the burden of dealing with conflicting expectations in society today. The purpose of this research study was to compare expectations of teachers and peers for gifted females and to discuss the implications of these expectations. It is expected that a significant difference exists between the expectations of teachers and peers for gifted female adolescents.

Method

Design

This study was a parallel-samples survey design, which is appropriate for studying attitudes and expectations. Data from two samples were collected simultaneously in order to conduct research related to the same research problem. Since data were collected at one point from each sample, this design is cross-sectional.

Subjects

The subjects comprising the two different samples of this study included teachers and peers from a southern, rural school setting. The teacher sample ($n = 28$) consisted of 11 male (39.3%) and 17 female (60.7%) teachers who taught both gifted and non-gifted students. The average age of the teachers was 32.3 years with a standard deviation of 5.83 years. Gifted students are defined as those who demonstrate abilities that give evidence of high performance in academic and intellectual aptitude. The years of teaching experience ranged from one to 20 years. The average teaching experience was 5.9 years with a standard deviation of 4.9 years. Of the teachers, 22 (78.6%) were classified as Caucasians, four (14.3%) were African-American, and two (7.1%) were Hispanic. The peer sample ($n = 34$) consisted of one seventh-grade and one eighth-grade randomly selected classroom of heterogeneously grouped students. There were 13 7th grade males, 9 8th grade males, 8 7th grade females, and 4 8th grade females. The students ranged in age from 11 to 14 years. The average age of the 7th grade sample was 12 years with a standard deviation of .54 years. The average age of the 8th graders was 13.61 years with a standard deviation of .50 years. Although a majority of the students were Caucasians (68%), the presence of a military base in the vicinity made the classrooms much more heterogeneous than other classrooms in the region. The students could be identified as African-American (20%), Hispanic (7%), and Asian and Pacific Islanders (5%). The peer group consisted of students who had not been classified as gifted.

Cluster sampling was the procedure used for selection of the peer sample in this study due to the availability of naturally grouped units of students in the classrooms.

Purposeful sampling was the procedure for the selection of the teacher sample, based on the researcher's identified criteria for sampling design as set forth by Kish (1965) and expanded in Wiersma, (1991): (a) goal orientation, (b) measurability, (c) practicality, and (d) economy. Cluster sampling and purposeful sampling provided the most practical and economical method of selection in order to gather the data required for this study.

Instrumentation

The measurement device used in this study was a 17 item, 4-point Likert-format forced-choice response questionnaire (ranging from strongly agree to strongly disagree) developed by the researcher to assess the attitudes and expectations of the two samples towards gifted female adolescents. The questionnaire (see Appendix) was based on various documented characteristics and behaviors of gifted female students identified by experts in the field of gifted education, counseling and psychology (Gassin, Kelly, & Feldhusen, 1993; Kerr, 1985; Kerr, 1991; Kravetz, 1984; Reis, 1987; Reis & Callahan, 1989). There was considerable overlap among the experts' opinions regarding the characteristics and behaviors of gifted female students. The seventeen items included in this questionnaire are reflective of recurring themes. The questionnaire was pilot tested with a sample from a similar school. Feedback from the pilot test was used to improve clarity of instructions and wording of the items. Local experts in Gifted Education and Educational Psychology provided a validity check for the items used. The improved clarity of directions and wording enhanced the reliability of the data (Wiersma, 1991). A forced-response Likert scale questionnaire was used which lends itself to statistical analysis.

Upon receipt of informed consent from parents and guardians at the beginning of the academic year, students were administered the questionnaire during an English class. The teachers received the questionnaire in their mailbox at school. They individually returned the completed questionnaires to the principal investigator.

Data Analysis

Following the collection of questionnaires from the two samples, coding of the data occurred. A Chi-square test was then used to determine if a relationship existed between any of the responses of the two groups. For purposes of the Chi-square analysis the four-point Likert scale responses were collapsed into two categories reflecting the percentage of agreement and disagreement. If no significant differences appeared between the two groups, then one could conclude that the responses representing the level of expectation for each item would be proportionately the same for each group. Expected responses

were computed for each sample of the 17 items on the questionnaire.

To determine whether the two groups' perceptions were similar or different, the following procedure was undertaken. Each group's agreement level (%) was converted to a ranking. Rankings for both groups were used to calculate Spearman's rank-order correlation.

Results

This study did support the hypothesis that a significant difference exists between expectations of teachers and peers. This proved true for gifted female adolescents on five of the 17 dimensions addressed. The teacher sample had the higher level of agreement of the two samples on 13 of the 17 questions. Teachers expected gifted females to make good grades, have large vocabularies, be creative, be original, be leaders, do well in math and science, be critical thinkers, be motivated, be flexible, be well-behaved, enjoy a challenge, be emotionally stable and be curious more than did the peer sample. The peer group indicated a higher level of agreement on only three of the 17 dimensions on the questionnaire. Peers held high agreement in being able to concentrate, having a good sense of humor, and being fun to be around. A Spearman rank-order correlation between the two groups' agreement level on the 17 dimensions was .62. This correlation was statistically significant ($p < .01$) indicating that there were more similarities than differences in the perceptions of the two groups. A coefficient of determination indicated that approximately 38.4% of the variation in the rankings of the teacher group could be explained by the variation in the rankings of the peer group.

There were significant differences among the responses of the two groups for five of the 17 items on the questionnaire. The critical value of Chi-square was determined to be 3.84 with the significance level of .05. Of the five differences that were significant, two were significant at the .01 level (critical value of Chi-square = 6.64). The significant differences in expectations (.05 level) occurred regarding good grades, leadership, and motivation, whereas differences in expectations regarding flexibility and curiosity were significant at the .01 level. These data are presented in Table 1.

The teacher sample (100%) agreed that gifted females should be curious, compared to 76.47% of the peer sample. Teachers (92.86%) surveyed also stressed flexibility, whereas only 55.88% of the peer sample agreed. Regarding leadership, 78.57% of the teachers felt gifted females should be leaders, compared to 47.06% of the peer sample. All of the teachers (100%) agreed that gifted females should be motivated, compared to 82.35 of the peer sample (see Table 2).

Table 1
Chi-square Results Concerning the Expectations
Placed on Gifted Girls by Teachers and Peers

Item	Chi-square	Significance
1. Good Grades	4.28	*
2. Large Vocabularies	3.82	<i>ns</i>
3. Creativity	3.80	<i>ns</i>
4. Originality	1.28	<i>ns</i>
5. Leadership	6.29	*
6. Math and Science Ability	1.05	<i>ns</i>
7. Critical Thinking	2.12	<i>ns</i>
8. Motivation	5.27	*
9. Flexibility	10.36	**
10. Concentration	0.50	<i>ns</i>
11. Well-Behaved	0.18	<i>ns</i>
12. Enjoy Challenge	2.05	<i>ns</i>
13. Sense of Humor	0.17	<i>ns</i>
14. Emotional Stability	1.61	<i>ns</i>
15. Popularity	0.01	<i>ns</i>
16. Curiosity	7.35	**
17. Fun to be Around	2.58	<i>ns</i>

Note. *Critical Chi-square = 3.84, $df = 1$, $p < .05$;

**Critical Chi-square = 6.64, $df = 1$, $p < .01$.

Table 2
Teacher and Peer Responses to Questionnaire

	Percent			
	Teacher ($n = 28$)		Peer ($n = 34$)	
Gifted girls should:	Agree	Disagree	Agree	Disagree
1. make good grades	100.00	0.00	85.29	14.71
2. have large vocabularies	82.14	17.86	58.82	41.18
3. be creative	96.43	3.57	79.41	20.25
4. be original	96.43	3.57	88.24	11.76
5. be leaders	78.57	21.43	47.06	52.94
6. do well in math and science	82.14	17.86	70.59	29.41
7. be critical thinkers	92.86	7.14	79.41	20.59
8. be motivated	100.00	0.00	82.35	17.65
9. be flexible	92.86	7.14	55.88	44.12
10. be able to concentrate	92.86	7.14	97.06	2.94
11. be well-behaved	89.29	10.71	85.29	14.71
12. enjoy a challenge	96.43	3.57	85.29	14.71
13. have a good sense of humor	71.43	28.57	76.47	23.53
14. be emotionally stable	82.14	17.86	67.65	32.35
15. be popular	25.00	75.00	26.47	73.53
16. be curious	100.00	0.00	76.47	23.53
17. be fun to be around	53.57	46.43	73.53	26.47

Discussion and Conclusions

The results of this study appear consistent with the available literature (Johnson & Lewman, 1990; Louis & Lewis, 1992; Luftig & Nichols, 1991) on expectations of gifted female adolescents. Based on the survey responses, it is clear that expectations for gifted females differ

between teachers and peers. Teachers have significantly higher expectations for gifted females in the areas of leadership, flexibility, and curiosity than peers. Whereas the adult expectations focused on performance, the peer sample had a higher regard for a characteristic allowing one to fit in with a group.

The literature (Clark, 1992; Clinkenbeard, 1991; Kleinsasser, 1986; Kline & Short, 1991; Luftig & Nichols, 1991; Reis & Callahan, 1989; Sisk, 1987) clearly indicates that gifted females are faced with conflicting expectations in society, be they teacher or peer expectations. The school (teachers) and peers are two of the primary socialization agents in society.

All of the teachers agreed that gifted females need to make good grades, be motivated, and be curious. Nongifted adolescent girls may be more inclined to focus their energies in the social arena. Therefore, gifted females might perform poorly in class or suppress their curiosity in favor of being accepted socially by their peers. This is one example of how conflicting expectations could influence a gifted female, especially in the adolescent years.

This study supported the suggestion that gifted females face a dilemma in today's society in dealing with conflicting societal expectations. Reis (1994) has pointed out that differences are found between adolescent boys and girls in their expectations about future education, career, and family; attitudes about school and school achievement; and concept of gender differences. Since gifted female adolescents differ from gifted adolescent males with respect to identity in the psychological and social dimensions (Leroux, 1986), these students need special guidance. As Debold (1995) states, "Girls need the support of adults to resist pressures to conform to outdated stereotypes that can limit their expectations and achievement" (p. 24). This is only possible when significant adults, who have influence in the lives of gifted females, are adequately educated about the detrimental effects of confusing, conflicting societal expectations.

References

- American Association of University Women (AAUW). (1991). *Shortchanging girls, shortchanging America: A nationwide poll to assess self-esteem, educational experiences, interest in math and science, and career aspirations of girls and boys ages 9-15*. Washington, DC: Author. (ERIC Document Reproduction Service No. ED 340 657).
- Angrist, S. S. (1969). The study of sex roles. *Journal of Social Issues*, 25, 215-232.
- Backes, J. S. (1994). Bridging the gender gap: Self-concept in the middle grades. *Schools in the Middle*, 3(3), 19-23.
- Castenell, L. A. (1983). Achievement motivations: An investigation of adolescent achievement pattern. *American Educational Research Journal*, 20, 503-510.
- Clark, B. (1992). *Growing up gifted: Developing the potential of children at home and at school* (4th ed.). New York: Macmillan.
- Clinkenbeard, P. (1991). Unfair expectations: A pilot study of middle school students' comparisons of gifted and regular classes. *Journal for the Education of the Gifted*, 15(1), 56-63.
- Cooley, D., Chauvin, J. C., & Karnes, F. A. (1984). Gifted females: A comparison of attitudes by male and female teachers. *Roeper Review*, 6(3), 164-167.
- Debold E. (1995). Helping girls survive the middle grades. *Principal*, 74(3), 22-24.
- Dickens, M. N., & Cornell, D. G. (1993). Parent influences on the mathematics self-concept of high ability adolescent girls. *Journal for the Education of the Gifted*, 17, 53-73.
- Dweck, C., & Bush, E. (1976). Sex differences in learned helplessness: Differential debilitation with peer and adult evaluators. *Developmental Psychology*, 12, 147-56.
- Erickson, E. H. (1950). *Childhood and society*. New York: W.W. Norton.
- Gassin, E. A., Kelly, K. R., & Feldhusen, J. F. (1993). Sex differences in the career development of gifted youth. *The School Counselor*, 41, 90-95.
- Gilligan, C. (1982). *In a different voice*. Cambridge: Harvard University Press.
- Gilligan, C., Lyons, N. P., & Hammer, T. J. (Eds.). (1990). *Making connections: The relational world of adolescent girls at Emma Willard School*. Cambridge, MA: Harvard University.
- Heller, K. A., & Ziegler, A. (1996). Gender differences in mathematics and the sciences: Can attributional retraining improve the performance of gifted females? *Gifted Child Quarterly*, 40, 200-209.
- Hingham, S., & Navarre, J. (1984). Gifted adolescent females require differential treatment. *Journal for the Education of the Gifted*, 8, 43-58.
- Horner, M. S. (1987). Toward an understanding of achievement related conflicts in women. In M. R. Walsh (Ed.), *The Psychology of Women* (pp. 169-184). New Haven: Yale University.
- Howard-Hamilton, M., & Robinson, T. L. (1991). Sex role characteristics of female governor's school students. *Gifted Child Today*, 14(4), 32-35.
- Johnson, L. J., & Lewman, B. S. (1990). Parent perceptions of the talents of young gifted boys and girls. *Journal for the Education of the Gifted*, 13(2), 176-188.
- Kerr, B. (1985). Smart girls gifted women: Special guidance concerns. *Roeper Review*, 8, 30-33.

- Kerr, B. (1991). Educating gifted girls. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 402-415). Boston: Allyn and Bacon.
- Kish, L. (1965). *Survey sampling*. New York: Wiley.
- Kleinsasser, A. M. (1986). *Exploration of an ambiguous culture: Conflicts facing gifted females in rural environments*. Paper presented at the Annual Conference of the National Rural and Small Schools Consortium. Bellingham, WA. (ERIC Document Reproduction Service No. ED 278 522).
- Kline, B. E., & Short, E. B. (1991). Changes in emotional resilience: Gifted adolescent females. *Roeper Review*, 13(3), 118-121.
- Kravetz, N. (1984, November). *Looking at the special needs of gifted girls*. Paper presented at the Annual Institute for the Gifted, Los Angeles, CA.
- Leroux, J. A. (1986, April). *Sex differences influencing gifted adolescents: An ethnographic study*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Leung, S. A., Conoley, C. W., & Scheel, M. J. (1994). The career and educational aspirations of gifted high school students: A retrospective study. *Journal of Counseling & Development*, 72, 298-303.
- Louis, B., & Lewis, M. (1992). Parental beliefs about giftedness in young children and their relation to actual ability level. *Gifted Child Quarterly*, 36(1), 27-31.
- Luftig, R. L., & Nichols, M. L. (1991). An assessment of the social status and perceived personality and school traits of gifted students by non-gifted peers. *Roeper Review*, 12(3), 148-152.
- Orenstein, P. (1994). *Schoolgirls: Young women, self-esteem, and the confidence gap*. New York: Doubleday.
- Randall, V. (1997). Gifted girls: What challenges do they face - A summary of the research. *Gifted Child Today*, 20, 42-49.
- Reis, S. M. (1994). Attitudes of adolescent gifted girls and boys toward education, achievement, and the future. *Gifted Education International*, 9(3), 144-151.
- Reis, S. M. (1987). We can't change what we don't recognize: Understanding the special needs of gifted females. *Gifted Child Quarterly*, 31(2), 83-89.
- Reis, S. M., & Callahan, C. M. (1989). Gifted females: They've come a long way - or have they? *Journal for the Education of the Gifted*, 12(2), 99-117.
- Sadker, M. P., & Sadker, D. M. (1986). Sexism in the classroom: From grade school to graduate school, *Phi Delta Kappan*, 67, 513.
- Sadker, M., & Sadker, D. (1994). *Failing at fairness: How America's schools cheat girls*. New York: Scribner's.
- Sadker, M., Sadker, D., & Klein, S. (1991). The issue of gender in elementary and secondary education. In G. Grant (Ed.), *Review of Educational Research* 17 (pp. 269-334), Washington, DC: American Educational Research Association.
- Sisk, D. A. (1987). *Teacher perceptions of gifted*. Address to the Hillsboro Association for Gifted, Tampa, FL.
- Wiersma, W. (1991). *Research Methods in Education* (5th ed.). Boston: Allyn and Bacon.

Appendix Questionnaire

Directions: Below are statements about characteristics, behaviors and attitudes describing the ideal gifted female adolescent. For each item, indicate the extent of your agreement, from Strongly Agree to Strongly Disagree, by checking the appropriate category.

Gifted Girls should:
make good grades.
have large vocabularies.
be creative.
be original.
be leaders.
do well in math and science.
be critical thinkers.
be motivated.
be flexible.
be able to concentrate.
be well-behaved.
enjoy a challenge.
have a good sense of humor.
be emotionally stable.
be popular.
be curious.
be fun to be around.

[illegible]

WISC-III Stability Over a Three-Year Period in Students with Learning Disabilities

Teresa Smith, Billy L. Smith, Ronald K. Bramlett, and Nancy Hicks
University of Central Arkansas

The stability of IQ scores has often been debated among school psychologists and special educators responsible for making special education placement decisions. This study examined the stability of the WISC-III over a three-year period. WISC-III data were collected on 54 rural school children with learning disabilities. Mean Full Scale IQ scores did not change significantly between the first and second measurement period. However, 20% of the students showed a significant change. Thirty-three percent of the sample showed significant changes on at least one of the three IQ scales (Verbal, Performance, or Full Scale). School psychologists and special educators need to be aware of these potential significant IQ score shifts in order to make the most informed decisions about children's placements and the needed evaluation data.

The WISC-III is one of the most common intelligence tests used in making decisions for special education (Stinnett, Havey, & Oehler-Stinnett, 1994; Wilson & Reschly, 1996). When children with learning problems are evaluated, the WISC-III often plays a prominent role in special education decisions. Likewise, the WISC-III is used to assist in decisions for children who are already receiving special education services when they are reevaluated every three years. The Individuals with Disabilities Act of 1997 (IDEA) revised the reevaluation process for students in special education to be more flexible and in some cases make the use of intelligence testing optional. Thus, if an evaluation team decides that intelligence testing is not necessary in the three-year reevaluation, it is not required to be given. With these revised standards for reevaluations in special education, it is critical to examine the stability of commonly used IQ tests such as the WISC-III in order to make the most informed decisions for students. While group IQ scores are relatively stable across time, IQ stability for individual children may vary greatly and is an important consideration that may have educationally significant consequences (Sameroff, Seifer,

Baldwin, & Baldwin, 1993). There is a need to document IQ shifts in populations such as children with learning disabilities as well as in the general population.

A number of studies were conducted on the WISC-R to measure stability over time (e.g., Bauman, 1991; Smith, 1978; Truscott, Narrett, & Smith, 1994; Vance, Blixt, Ellis, & Debell, 1981). Vance et al. (1981) found high correlations between the scales of the WISC-R covering a period of 2 years, 2 months. They found correlations of .80 for Verbal IQ, .91 for Performance IQ, and .88 for Full Scale IQ. They concluded that IQ scores obtained by the WISC-R were stable for students with learning problems.

Other researchers (Bauman, 1991; Truscott et al., 1994) studying the WISC-R have found correlations to be lower. For example, Truscott et al. found the correlations to be .69, .72 and .77 respectively for the Verbal IQ, Performance IQ, and Full Scale IQ for the same students evaluated 3 years apart. Truscott et al. also considered frequency count data as a measure of IQ stability. They found, that although moderate correlations were obtained, the sample showed substantial numbers of students with large shifts in IQ (13% showing more than 10 points change). Using the WISC-R SEM of 3.19 points, Bauman found that 58% of his sample deviated more than 1 SEM from their first to second score, 40% more than 2 SEM, 22% differed by more than 3 SEM, 10% by more than 4 SEM, and 3% deviated more than 5 SEM. He pointed out that these differences should be considered in making interpretations about the results. Bauman also found age at testing to be relevant, with greater changes in IQ scores in younger children than with older children. In children with higher IQs, he found larger drops in IQ. Therefore, he suggested consideration about retesting be given to the age of the child and the range of the IQ.

Teresa Smith received her Ph.D. in Psychology from the University of Mississippi and is an Associate Professor of Psychology at the University of Central Arkansas. Billy L. Smith received his Ph.D. in Psychology from the University of Mississippi and is a Professor of Psychology at the University of Central Arkansas. Ronald K. Bramlett received his Ph.D. in Psychology from the University of Cincinnati and is an Associate Professor at the University of Central Arkansas. Nancy Hicks is a graduate student in the School Psychology Program at the University of Central Arkansas. Please address all correspondence regarding this article to Teresa D. Smith, Box 4915 UCA, Conway, AR, 72035-0001 or by e-mail to Terrys@mail.uca.edu.

A few studies have been conducted with the WISC-III (Wechsler, 1991). Slate, Jones and Saarnio (1997) in a study of the WISC-III stability found no significant differences in the initial and three-year reevaluations on Verbal, Performance, and Full Scale IQ scores. They reported that test-retest scores were significantly correlated. They had a relatively small sample of 22 children with learning disabilities and 12 children who were mentally retarded. Individual changes were not reported. Stavrou and Flanagan (1996) studied the stability of WISC-III scores in a sample of 50 children with learning disabilities. Very little change was found over-time in the mean Full Scale IQs. However, 14% of their participants changed a significant nine or more points. On the Verbal IQ measures, the mean scores did not change significantly. However, 14% of the participants changed 10 or more points in either direction. Finally, on the Performance IQ test, the mean scores did not change significantly. Compared to the Verbal and Full Scale IQ scores, a higher percentage (24%) of students showed changes of 13 or more points. Changes in positive and negative directions tended to keep the averages about the same. Stavrou and Flanagan stated that clinically significant changes occurred in some individuals and this phenomenon should be considered during reevaluations. Reasons for these changes need to be determined and taken into account in the students' programming.

Slate and Wickes (1998) compared WISC-III evaluations over a three-year period with 49 special education students (20 learning disabled, 23 mentally retarded, and 6 who did not meet special education eligibility requirements at the time of reevaluation). They found correlations of .83 for Full Scale IQ, .88 for Verbal IQ, and .78 for Performance IQ. Although high correlations were found, drops in Full Scale, Verbal, and Performance IQs were, on average, 8.1, 7.9, and 6.9. Although they found strong evidence of WISC-III stability between the 3-year testing period, these researchers did not include the amount of IQ decrease for individual children. However, they did find that students whose IQs were higher at initial testing dropped more than students with lower IQs, possibly due to regression to the mean. Slate and Wickes recommended that researchers further investigate the generalizability of their findings including the drops in IQs for children in special education.

It is clear from previous research that the WISC-III is a stable instrument across time for groups of children. Yet, significant drops are likely to occur for children in special education. In order to examine further the stability of the WISC-III in students with learning disabilities, the present study examined the relationship between two administrations of the WISC-III at 3-year intervals and the degree of IQ change for individual children.

Method

Participants

The participants in this study were 54 (grade range was from 1st - 8th) school-aged children who had been classified as learning disabled (LD) in accordance with Arkansas' Program Standards and Eligibility Requirements for Special Education (1993). Thirty-two were males, and 22 were females, with 45 Caucasian and 9 Blacks. Subjects were drawn from two rural schools in central Arkansas and were selected based on the availability of two separate administrations of the WISC-III with an average of 3 years between administrations. The modal grade level on the first test administration was 3rd grade, and the modal grade level for the second test was 6th grade. The mean age at the first administration for the total sample was 9.96 ($SD = 2.14$) years, whereas the mean age for the second test was 12.96 ($SD = 2.34$) years. The WISC-III tests were administered by certified school psychologists.

Instrument

The WISC-III is the most widely used intelligence test by school psychologists (Wilson & Reschly, 1996). The test was standardized on a representative sample of the U.S. population of children. The sample consisted of 2,200 cases including 200 children in each of 11 age groups ranging from 6 through 16 years. Variables such as gender, race/ethnicity, geographic region, and parent education were reflective of the U.S. population.

The WISC-III is comprised of 13 subtests (3 optional). Five subtests are included in the Verbal IQ scale and 5 are included in the Performance IQ scale. Both scales combined form the Full Scale IQ. Reliability for the WISC-III subtests range from $r = .69$ (object assembly) to $r = .87$ (vocabulary and block design). The reliability coefficients for the Verbal, Performance, and Full Scale IQs are .95, .91, and .96 respectively. Numerous validity studies have been conducted with the WISC-III and reported in the manual.

Procedures

Data were collected on IQ scores and subtest scores from the special education files for two school districts. Information on age, sex, race and grade was also obtained. Bonferroni-corrected paired t -tests between the IQ indices and subtests for each administration were calculated. Bonferroni-corrected correlations were also obtained between the first and second measures.

Results

Differences between the first and second administration of the WISC-III were calculated for the Verbal,

WISC-III STABILITY OVER A THREE YEAR PERIOD

Performance, and Full Scale IQs as well as for all the subtests excluding the supplementary tests. Pearson correlations were also calculated between the first and second administrations of the test. Table 1 presents the means, standard deviations, effect sizes, test-retest reliability coefficients, and *t*-test results of mean differences for FSIQ, VIQ, and PIQ.

The standard error of the difference between scores (Cohen, Swerdlik, & Phillips, 1996) was calculated for the FSIQ, VIQ, and PIQ. Applying Bonferroni-correction, these statistics were then multiplied by 2.41 to determine a value by which individual scores would need to differ in order to determine a significant shift. When rounded to the nearest whole number, the results indicated a value of 10 for Full Scale IQ, 11 for Verbal IQ, and 15 for Performance IQ. Cohen et al. indicated that this procedure is an appropriate way to determine a significant change for individuals. Individuals' difference scores were then compared to these values to determine how many of them had changed significantly. The percentage of students who changed significantly on at least one scale was 33 percent. Table 2 presents the percentage of individuals who changed significantly on the retest. Table 3 shows the dispersions of changes in IQ scores between the two testings.

Full Scale IQ. FSIQs did not change significantly as measured by the *t*-test (Table 1) although there was a decline of 2.06 points. The correlation between first and

second measure was .87. However, when considering change at the individual level, 20 percent of students showed a significant change in their IQ scores from the first to second test (Table 2). Seventeen percent of the whole sample demonstrated a significant decline on the retest.

Verbal IQ. Mean VIQ scores declined 3.54 points which is a significant change as measured by the *t*-test (Table 1). The correlation between first and second measures on VIQ was .83. When the number of individuals making a significant change was examined, 19% scored above the cut-off of 11 points difference. Seventeen percent were found to show a significant drop in VIQ.

Performance IQ. PIQs did not change significantly as measured by the *t*-test (Table 1), although a gain of .13 points was present. However, as shown in Table 2, 15 percent of students showed a significant change in their IQ score from the first to second test. Seven percent of the overall sample showed a significant decline in PIQ.

Subtest results. Correlations between first and second testings were significant on all subtests except Picture Completion. Significant declines in average scores were noted on Arithmetic and Vocabulary. Decline in IQ scores were more common on Verbal subtests than Performance subtests, and these two subtests may reflect lack of success in school. Overall, it appears that the group analysis indicated stable scores in most subtests.

Table 1
Means, Standard Deviations, Correlations (*r*), and *t* Values (*t*).

	Mean		SD		D**	<i>r</i>	<i>t</i>
	Test	Retest	Test	Retest			
Verbal IQ	90.28	86.74	13.40	12.82	.27	.83*	3.40*
Performance IQ	92.39	92.52	14.02	14.36	.01	.78*	-.10
Full Scale IQ	90.43	88.37	12.87	12.74	.16	.87*	2.30
Information	8.08	7.89	2.59	2.96	.07	.73*	.34
Similarities	8.39	8.06	3.07	2.54	.12	.63*	.79
Arithmetic	7.98	6.70	3.04	2.44	.47	.46*	2.98*
Vocabulary	8.06	7.24	3.05	2.53	.29	.83*	3.27*
Comprehension	8.79	8.22	3.11	3.37	.18	.49*	.82
Picture Completion	8.79	9.15	2.92	3.32	.12	.37	.66
Coding	9.41	8.48	3.20	3.45	.28	.50*	.54
Picture Arrangement	8.98	9.20	2.99	3.35	.07	.60*	-.48
Block Design	8.25	8.24	3.59	3.73	.00	.78*	-.11
Object Assembly	7.85	8.61	3.17	2.80	.25	.58°	-2.10

Note: * indicates $p < .05$ with a Bonferroni correction, °° denotes effect size (the difference between the mean of test 1 and the mean of test 2 divided by the average standard deviation of the two tests).

Table 2
Percentage of Students Demonstrating Statistically Significant
($p < .05$) Changes in FSIQ, VIQ, and PIQ

Scale	Criterion	Percent
FSIQ	≥ 10 points	20
VIQ	≥ 11 points	19
PIQ	≥ 15 points	15

Table 3
Frequency of Changes (Plus or Minus) for IQS
Over a Three-Year Test-Retest Interval

WISC-III Scales	Score Change									
	0-3	4-6	7-9	10-12	13-15	16-18	19-21	22-24	25-27	
Verbal IQ	20	13	8	4	5	2	2	0	0	
Performance IQ	12	18	8	5	4	6	0	0	1	
Full Scale IQ	25	10	8	7	3	1	0	0	0	

Discussion

When using the measures of mean change and correlations, the WISC-III was relatively stable for this group of students with learning disabilities. Even the significant change occurring in VIQ was only 3.54 points. The effect sizes ranged from .00 (Block Design) to .47 (Arithmetic). These effect sizes are generally considered small to medium (Cohen, 1996). The data obtained were in general agreement with other research in this area (e.g., Bauman, 1991; Slate et al. 1997; Stravou & Flanagan, 1996; Truscott et al., 1994). As in similar studies (e.g., Bauman; Stravou & Flanagan), significant drops in IQs were found for individual students with learning disabilities. The overall drop in IQs from about 90 to 88 seems counterintuitive when considering regression effects. One would expect the IQ scores to move closer to the mean of 100, however, this was not the case. This sample was not different from other samples of LD students in which the range in IQs is around 90 - 93 (Bender, 1998).

The patterns of IQ drops were especially prevalent in verbal subtests (e.g., Arithmetic and Vocabulary). These subtests are most likely influenced by academics. One plausible explanation for these drops may be related to the modified curriculum to which students with learning disabilities are exposed. Some have argued that the curriculum in special education is less rigorous and special education students work at a slower pace which may explain in part why these decreases are seen (e.g., Stainback & Stainback, 1984).

Also, students with learning disabilities often have language deficits that make it difficult to learn new information and to pull stored information from memory (e.g., Smith, 1998). Language deficits may also affect reading comprehension and written expression. Students with higher verbal skills tend to improve more rapidly in reading than students with lower verbal skills (Kuder, 1991).

The present study has limitations in that only children with learning disability placements were used. The data also come from two rural school districts. Further research should be conducted on other populations, particularly examining the individual changes in IQ scores and their impact on the definition of learning problems.

Practice Considerations

School evaluation teams should consider the possibility of large differences in IQs including drops in IQ between evaluations. Many states use intelligence/achievement discrepancies including regression analyses to assist in classifying students as learning disabled. With substantial percentages of students showing declines in FSIQ and VIQ, a student may no longer qualify by regression analysis upon a triennial reevaluation. Teams should consider this effect in a reevaluation in order to make the most informed decisions for students. Certainly, in cases where significant drops are observed, teams need to examine the possible reasons and how those changes might potentially impact special education placement and planning decisions.

Care should be taken to utilize multiple sources of data related to students' functioning when considering placement decisions and especially for programmatic planning and IEP development. Procedures such as curriculum-based assessment as directed by IDEA (1997) should be used to supplement intelligence testing to assist school psychologists and special educators in planning the most effective academic interventions for students.

References

- Bauman, E. (1991). Stability of WISC-R scores in children with learning difficulties. *Psychology in the Schools*, 28, 95-100.
- Bender, W. N. (1998). *Learning disabilities: Characteristics, identification, and teaching strategies* (3rd ed.). Needham Heights, MA: Allyn and Bacon.
- Cohen, B. H. (1996). *Explaining psychological statistics*. Pacific Grove, CA: Brooks/Cole.
- Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment* (3rd ed.). Mountain View, CA: Mayfield Publishing Company.

WISC-III STABILITY OVER A THREE YEAR PERIOD

- Kuder, S. J. (1991). Language abilities and progress in a direct instruction reading program for students with learning disabilities. *Journal of Learning Disabilities*, 24, 124-127.
- Program Standards and Eligibility Criteria for Special Education. (1993). Little Rock, AR: Department of Education.
- Public Law 105-17. *Individuals with Disabilities Education Act Amendments of 1997*.
- Sameroff, A. J., Seifer, R., Baldwin, A., & Baldwin, C. (1993). Stability of intelligence from preschool to adolescence. *Child Development*, 64, 80-97.
- Slate, J. R., Jones, C. H., & Saarnio, D. A. (1997). WISC-III IQ scores and special education diagnosis. *The Journal of Psychology*, 13, 119-120.
- Slate, J. R., & Wickes, K. (1998). Longitudinal stability of WISC-III IQs. *Journal of Special Education*, 21, 74-80.
- Smith, C. R. (1998). *Learning Disabilities: The Interaction of Learner, Task, and Setting* (4th ed.). Boston: Allyn and Bacon.
- Smith, M. D. (1978). Stability of WISC-R subtest profiles for learning disabled children. *Psychology in the Schools*, 15, 4-7.
- Stainback, W., & Stainback, S. (1984). A rationale for the merger of special and regular education. *Exceptional Children*, 51, 102-111.
- Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment*, 12, 331-350.
- Stravou, E. & Flanagan, R. (1996, March). *The stability of WISC-III scores in learning disabled children*. Paper presented at the National Association of School Psychologists Annual Convention, Atlanta, GA.
- Truscott, S. D., Narrett, C. M., & Smith, S. E. (1994). WISC-R subtest reliability over time: Implications for practice and research. *Psychological Reports*, 74, 147-156.
- Vance, H. B., Blixt, S., Ellis, R. & Debell, S. (1981). Stability of the WISC-R for a sample of Exceptional Children. *Journal of Clinical Psychology*, 37, 397-399.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-III*. San Antonio: The Psychological Corporation.
- Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review*, 25, 9-23.

Cooperative Editing: An Effective Strategy for Error Reduction in Remedial Freshman Composition

Sue S. Minchew and Neil G. Amos
Mississippi State University

The purpose of this study was to determine whether cooperative editing would significantly decrease the mean number of major (nonstandard) grammatical errors on papers written by remedial freshman composition students. Four intact remedial freshman composition classes (79 students) taught by the senior author were randomly labeled experimental or control. All four classes were taught such aspects of composition as organization, development, word choice, and sentence structure. However, the two control classes were taught standard written grammar by the textbook/lecture/discussion method in contrast to the experimental classes, which were taught a writer's grammar through researcher-developed cooperative editing exercises. A pretest paper graded blindly by the senior author served as a covariate to a posttest paper graded by the same process. Using the mean number of major grammatical errors per group as the outcome measure, an ANCOVA revealed significantly fewer ($p < .05$) major grammatical errors on posttest papers written by experimental students than on those written by control students. Results verified the efficacy of using cooperative editing to teach the application of standard written grammar.

Despite the writing deficiencies of many beginning college students, particularly in the use of an acceptable written grammar, some researchers suggest that grammar instruction has no place in composition courses, that the study of grammar indeed detracts from the writing process by placing too much emphasis on correctness at the expense of effective communication (Bennett, Jordan, Long, & Wade, 1976; Braddock, Lloyd-Jones, & Shoer, 1963; Hartwell, 1985; Holden, 1994). However, other researchers disagree, defending grammar instruction in writing classes (Claywell, 1995; Delpit, 1995; Hairston, 1981; McCleary, 1995; Neulieb & Brosnahan, 1987; Olson, 1995; Shaughnessy, 1977; Vavra, 1985). Formal grammar instruction, which Holden (1994) described as lecture/discussion of grammatical principles and definitions accompanied by examples and exercises, has been most maligned. Noguchi (1991) contended that formal grammar instruction has resulted in negligible improvement in student writing primarily because grammar has been taught the wrong way, geared neither to writing nor to the students.

Recent scholarship indicates a resurgence of interest in teaching grammar in writing classes. Noting that the return of grammar teaching has resulted from the general dismay over writing assessment and concerns about errors

in student writing, McCleary (1995) mentioned three needs in addressing the problem: (a) more accurate rules of correctness, (b) a more teachable grammar that students can apply, and (c) a better pedagogy than the workbook. Several writing teachers (e.g., Noguchi, 1991; Shaughnessy, 1977; Weaver, 1996) agreed that grammar does have its place in writing instruction. However, instruction should not include formal grammar but should be related to a student's own writing (Connors & Lunsford, 1988; Noguchi; Renwick, 1994; Rose, 1996; Shaughnessy; Weaver). Noguchi advocated teaching less grammar and concentrating more on those basic rules relevant to the writing process.

Although some composition teachers have begun to ignore "superficial errors" in an attempt to de-stigmatize the language of a student's community, several scholars (e.g., Claywell, 1995; Delpit, 1995; Olson, 1995) claim that it is a disservice, especially to nontraditional students, not to provide instruction in grammar if that is what they need. Shaughnessy (1977) argued that grammar study cannot be ignored at the basic writing level because of the relationship between writer and audience, a relationship onto which errors intrude. Shaughnessy further claimed that attempts to minimize or excuse students' grammatical problems only cause confusion. Claywell noted that the educated public, for whom the students will eventually write, remains very critical of grammatical flaws. Hairston's (1981) extensive study, which surveyed professional colleagues outside of English about their reactions to types of errors in written work, verified the importance of correctness to professionals. Hairston discovered that the errors regarded as most offensive by

Sue S. Minchew is an Associate Professor and Neil G. Amos is Professor Emeritus of Curriculum and Instruction at Mississippi State University. The authors wish to extend a special thanks to Christopher Skinner for his help in the revision of the original manuscript. Correspondence regarding this article should be directed to Sue S. Minchew, Box 9705, Mississippi State, MS 39762 or by e-mail to minchew@ra.msstate.edu.

the respondents included obvious verb tense or agreement errors, fragments, comma splices, and object pronouns used as subjects. Shaughnessy extolled the importance of training students in their perceptions of errors so that by stages they can learn to proofread their own papers. However, like Meyer (1986), Shaughnessy contended that traditional grammar-teaching was not the answer.

The answer may lie in cooperative learning. It has been well documented that lower level and minority students benefit academically from working in a non-competitive group atmosphere (Cuban, 1989; Davidson & Koppenhaver, 1993; Slavin & DeVries, 1979). Sagor (1993) posited that at-risk students tend to be more highly motivated by tasks that seem useful, that seem to benefit them personally, than they are by seemingly irrelevant workbook or textbook exercises. Cooperative work helps such students to work as a team and to develop greater feelings of usefulness. Sagor also noted that at-risk students in particular view school as a social experience and that cooperative learning builds on that perception. Several researchers have found that cooperative learning has positive effects on student achievement and attitude (Johnson, Johnson, & Smith, 1991; Presseisen, 1992; Slavin, 1980; Slavin & DeVries, 1979). Slavin's review of studies on cooperative learning revealed that 90% of the studies resulted in increased achievement when cooperative learning was used as the intervention. In his meta review of studies on various teaching modes, Hillocks (1986) found that the environmental mode (minimal lecture, small group tasks to prepare for later individual work) produced superior achievement when compared to the presentational mode (teacher explanation, individual work).

Several writing teachers have found content improvements in student papers as a result of reader response groups, a collaborative group process in which students judge the content of each other's rough drafts (Elbow, 1973; Gebhardt, 1980; Thompson, 1988). Some improvements in grammar and mechanics have resulted from peer editing, which consists of group members proofreading each other's papers (Beachy, 1992; Herreman, 1988).

Although much research on cooperative learning exists, some gaps remain that still need to be investigated. Yager, Johnson, and Johnson (1985), for example, pointed out the need for studies that examine both immediate and long-term effects of collaborative learning on student achievement. Talmage, Pascarella, and Ford (1984) reported that 75% of the studies on cooperative learning covered a period of 10 weeks or less. Bruffee (1993) claimed that the college classroom continues to be dominated by competitive and individualistic learning in spite of what research has shown to be effective (Johnson et al., 1991). Moreover, because most of the studies which examined the effects of cooperative learning on at-risk

students used heterogeneous groups of high, middle, and low achievers, a need exists for studies which examine the effects of cooperative learning on low achievers working in homogeneous groups. Whereas several studies have shown increased achievement in writing classes using cooperative learning at the elementary, middle school, and secondary levels, very few studies have focused on cooperative learning in college writing classes (Johnson et al., 1985). One small study found significant improvements in the writing of freshman students taught by cooperative learning methods (Hart, 1991). Another researcher described more positive attitudes toward writing among his freshman students as a result of cooperative learning (Ney, 1989).

The current study incorporated elements of methods that other researchers had already found to be effective for teaching writing and grammar. It involved a process approach (Holden, 1994); it utilized the students' own writing (Hunter & Wallace, 1995; Hurst, 1987); it geared the study of grammar to the writing process (Noguchi, 1991; Shaughnessy, 1977; Vavra, 1985); it used a non-traditional teaching method (Meyer, 1986; Shaughnessy, 1977); and it utilized the environmental mode (Hillocks, 1986) with cooperative learning (Donovan, 1990; Hart, 1991), an appropriate teaching strategy for at-risk writing students (Davidson & Koppenhaver, 1993).

In developing a pedagogy for the remedial composition students in the current study, the researchers focused specifically on the students' deficiencies in standard grammar in devising a teaching strategy that integrated grammar with the writing process. They did not want grammar *per se* to become the only focus of the course; it was, after all, a writing course. Nor did they want to abandon totally the traditional teaching strategies (lecture, discussion, textbook, reader response groups) which had been somewhat successful for a number of years. Aware of the importance of peer group acceptance, of research studies documenting increased achievement for at-risk students taught by nontraditional methods, and of the grammatical needs of remedial writing students, the researchers developed a strategy they have termed cooperative editing (a type of cooperative learning) as a means of training remedial students to understand, find, and correct nonstandard grammatical errors in their own writing. No previous researchers have examined the effects of cooperative editing on writing classes.

The purpose of the current study was to evaluate and compare the effects of two methods of composition instruction on remedial college composition students' written grammar. Both control and experimental groups were taught such aspects of composition as organization, development, word choice, and sentence structure. However, the control group was taught standard written grammar by the textbook/lecture/discussion method in

contrast to the experimental group, which was taught a writer's grammar through cooperative editing exercises. For these exercises, students were randomly assigned to groups to analyze and correct anonymous student papers containing major grammatical errors. (See Appendix C for a list and definitions of these errors).

Method

Participants

The sample consisted of 79 students in four intact Basic English classes (remedial) at a medium-sized southern university. Fourteen students were lost or excluded from the original 93. Five students dropped, four quit attending, and five were eliminated because they did not participate in 80% of the editing activities, a requirement for inclusion in the study. In all, 33 females, 46 males, 45 African Americans, 32 Caucasians, and 2 Native Americans participated in the study. The two experimental classes included 38 students, 19 in each class, with the following demographics: 13 females, 25 males, 23 African Americans, 13 Caucasians, and 2 Native Americans. Among the control students (classes of 20 and 21) were 20 females, 21 males, 22 African Americans, and 19 Caucasians. The average age of the sample was 19.8 ($SD = 4.2$); the average age of experimental students was 20.4 ($SD = 4.5$), and of the control students 19.2 ($SD = 3.8$).

Prior to the first class meeting, the four classes were randomly assigned to experimental or control conditions. In an effort to minimize the effects of class times, experimental and control groups each included one morning and one afternoon class.

Initially assigned to remedial composition classes on the basis of ACT English scores between 15 and 18 (none had SAT scores), the students were homogenous in terms of ability. According to data obtained from the Office of Institutional Research (1998), the mean composite ACT score for the sample was 17.4 ($SD = 1.5$); for the experimental students, it was 17.3 ($SD = 1.4$); and for the control students, it was 17.5 ($SD = 1.6$). An analysis of variance conducted on the pretest papers verified that no statistically significant *a priori* differences existed among the four classes: $F(3, 78) = 1.27, p > .05$. A *t*-test comparing the combined experimental classes with the combined control classes on the pretest paper also yielded nonsignificant results ($t = .97, p > .05$).

Design

A quasi-experimental pretest posttest control group design was used for the study (Borg & Gall, 1989). As Borg and Gall advocated for experiments involving small, intact classes, the two cooperative editing classes were

combined to form the experimental group, and the two traditionally taught classes were combined to constitute the control group. This combination resulted in larger *ns* and, therefore, greater power. It also kept the focus of the study on the treatment rather than on individual classes.

Instrumentation

The dependent variables included a pretest paper and a longer posttest paper; for each, the outcome measure was the average number of major grammatical errors committed on the papers. Grades on papers were not considered. The pretest paper was a 100-150 word paper written during the second class meeting on a previously unannounced topic ("Why I want to be _____"). Students were instructed to record their names only on cover sheets to their papers. All four sets of papers were subsequently delivered to an English Department secretary who number-coded each paper and cover sheet, removed the cover sheets, and shuffled the papers before returning them. The papers were then graded blindly by the senior author, who had knowledge neither of the identities of the students nor of the class to which they belonged.

Approximately 300 words in length, the posttest paper was written entirely in class during two class meetings near the end of the semester. Graded by the same blind process as the pretest paper, it was used to determine whether the students in the experimental classes made statistically significant gains in reducing the average number of major grammatical errors on their papers when compared to the average number of major grammatical errors on papers of the control students. A longer paper was used for this assessment since the additional opportunity for error would provide a clearer distinction between the two groups in terms of error reduction than a shorter paper would provide (Connors & Lunsford, 1988). The papers were written in class to ensure that they represented the students' own work; they were written near the end of the semester to maximize treatment time.

Analysis of Data

An Analysis of Covariance (ANCOVA) was used to test for differences in the average number of nonstandard errors made by experimental and control students on the posttest paper. The pretest paper served as the covariate. Even though the two groups of students had no significant differences in their ACT or SAT English scores nor any significant differences in the pretest papers, the researchers thought that the ANCOVA would provide the truest assessment of results comparing the performance of the two groups because it would adjust for any pre-existing individual differences.

Teaching Methods

The senior author was the instructor for all four classes, which were taught the same content with similar teaching methods except for the use of editing exercises in the experimental classes. Both control and experimental classes used *The Little Brown Handbook* (Fowler, Aaron, & Tavers, 1992) and *English Skills with Readings* (Langan, 1991) as textbooks. Detailed day-by-day syllabi and a daily journal helped ensure that class content and teaching methods remained the same for all four classes except for the treatment, cooperative editing. As prescribed by the departmental course description, papers written during the first half of the semester were 150-200 words in length; those written in the latter half were approximately 300 words in length. All of the papers were written entirely in class. Paragraphs or essays from Langan served as rhetorical models (description, narration, definition, exemplification, classification, and cause/effect) for all of the papers except the pretest. Lecture, discussion, and textbook readings were used in all four classes to teach such aspects of composition as organization of content, thesis sentence, outlining, paragraph construction and development, word choice, and sentence structure. All four classes participated in reader response groups, consisting of three or four students who critiqued the content of each others' papers using teacher-provided questions appropriate for the particular rhetorical model. Response groups were instructed to concentrate only on effectiveness of content, not on grammar.

Because of the students' severe grammatical deficiencies, grammar instruction was an essential component of all four classes. However, it was not formal grammar instruction as defined by Holden (1994) since no time was spent teaching definitions and identification of the parts of speech or parts of the sentence. Grammar, rather, was related to the writing process with the objective of eliminating nonstandard errors in the students' own writing. Although textbooks, lecture, and discussion were used in all four classes to introduce grammatical principles, the time spent introducing and explaining principles was minimized in the experimental classes. The Langan (1991) text contained especially helpful verb sections, which emphasized the differences between standard English and community dialect. For all four classes, an attempt was made to simplify grammatical concepts by avoiding difficult terminology and focusing primarily on the major nonstandard errors, what Noguchi (1991) called a writer's grammar. During the final editing stage of writing, all four classes used author-developed proofreading sheets (see Appendix A), which were divided into three sections: (a) verb errors (tense, subject-verb agreement), (b) sentence sense errors (comma splice, fragment, fused sentence), and (c) pronoun/possessive noun errors (pronoun-antecedent agreement, pronoun reference, case).

The sheets provided students (both experimental and control) with a method for proofreading that required them to analyze their own papers systematically for each type of major grammatical error, thus enabling them to focus on the grammar of their papers rather than on the content during the final proofreading/editing stage of composing. The amount of time spent on the study of grammar was approximately the same for all four classes. The main difference in the control and experimental classes was in the way they discussed and applied the rules of standard written grammar. More time was spent in lecture and whole class discussion in the control classes in contrast to more small group discussion in the experimental classes. Control students worked individually on textbook drills and exercises, whereas the experimental students practiced grammar application via cooperative editing exercises.

Developed from student papers written in the control classes, cooperative editing exercises were typed and given to the students as written except for slight manipulations of the grammar so that only major grammatical errors appeared in the papers. Occasionally a few errors were added so that a paper would contain at least 10 major errors. The experimental classes completed 11 editing exercises, 2 individually and 9 in cooperative editing groups. On each cooperative editing exercise (see Appendix B), students in groups of three or four were directed to correct the major grammatical errors contained in the student paper as well as to identify each error by name and corresponding rule number in *The Little Brown Handbook* (Fowler et al., 1992). To facilitate error recognition and the citation of corresponding rules, during editing work students could use their textbooks, notes, and researcher-developed Rules for Editing sheets (see Appendix C). These sheets listed and defined the major grammatical errors and provided section and page numbers in *The Little Brown Handbook* (Fowler et al.) for the related rules. After completing a cooperative editing exercise, each group submitted one copy, signed and proofread by all group members, to be graded. All members of a group received the same grade for the exercise. The first editing exercise contained only verb errors with each subsequent exercise incrementally adding the errors that had been studied since the previous one. Treated as minor quizzes, the 11 editing exercises were averaged together (with one drop grade) to constitute one grade equivalent to a paper grade.

Because it was difficult to ascertain students' abilities early in the semester and thus to select the recommended mixture of high, middle, and low ability students for groups (Johnson, Johnson, & Holubec, 1993; Slavin, 1987), group membership for cooperative editing was determined by a random drawing of numbers. After each of the two individual exercises, the students who made the

six highest grades were designated leaders of six different groups on the subsequent cooperative editing exercise.

Before the first cooperative editing exercise, students were instructed to work toward consensus in solving the grammatical problems found in the paper. Although each student in a group was given a copy of the exercise, students were encouraged to discuss the exercise paper line by line to determine errors, corrections, and rules. They were instructed to resolve differences by referring to their Rules for Editing sheets, their textbooks, or notes. For each exercise, group members were told to choose a reader and a recorder. Once the group reached consensus, the recorder transcribed the results onto an additional exercise sheet provided each group. Group members were advised to proofread this transcription carefully before signing the sheet to be sure that it accurately recorded the error names, corrections, and corresponding rules as agreed on by the group as their signatures indicated agreement with the exercise submitted for a grade. Students were informed that if they were unable to agree with other members of their group, they could submit their own individual "protest" exercise. (Interestingly, no such "protest" was ever submitted.)

After the initial instructions and number-drawing procedures, the instructor became a facilitator during cooperative editing, walking from group to group, occasionally asking questions and providing hints, but not answers. Although group participation was monitored via a brief questionnaire administered after the second cooperative editing exercise, the questionnaire indicated only one student who did not contribute sufficiently to group work. It seemed unnecessary thereafter to administer the questionnaire, as students began to work out the dynamics of group participation among themselves. Moreover, peer ethics of acceptance, familiarity, and reliability seemed to be more effective methods for ensuring and monitoring group participation than the questionnaire. Nonetheless, the questionnaire served to remind the students of the importance of participation and cooperation.

To facilitate accurate record-keeping and to document trends in the commission of major grammatical errors in all four classes, the number and kinds of major errors committed on each in-class paper were recorded on class error charts as well as on student papers with subnumbers added to letter grades to indicate the number of major errors contained in papers. Not only did the class error charts indicate individual class progress but also individual student progress toward the elimination of major errors. In addition, these charts were used to determine which errors needed to be retaught and, in the experimental classes, emphasized on subsequent editing exercises. Thus the students' own grammatical problems could be identified and addressed as advocated by

Noguchi (1991), Renwick (1994), and Rose (1996). To facilitate self-assessment, students in all four classes were required to keep similar individual error charts documenting their performance on each of their own papers (see Appendix D for class and individual student error charts). The individual charts provided students with a means of evaluating their own tendencies and progress over a number of papers. The charts also gave the students in all four classes direction in restudying their own errors.

Results

The results of the ANCOVA revealed that the experimental classes (38 students) committed statistically significantly fewer major grammatical errors on the post-test paper than the control classes (41 students) when the average numbers of major grammatical errors were compared: $F(1,76) = 4.17, p < .05$. The ANCOVA yielded an Eta squared effect size of .05. Nonsignificant results on the Bartlett-Box test ($p > .05$) and on the Shapiro-Wilks test ($p > .05$) indicated that the homogeneity of variance and normality assumptions had not been violated. The combined experimental classes had an adjusted mean of 1.76 ($SD = 1.95$), the combined control classes a mean of 2.76 ($SD = 2.30$).

Discussion

As expected, the cooperative editing students experienced a significantly greater reduction in the incidence of nonstandard errors on their own papers than did the traditional students. Research suggests that several differences in the types of instruction may account for these findings. First, cooperative editing exercises differed from textbook exercises in several significant ways: (a) They were developed from actual student papers and could thus be perceived as real and useful, factors according to Sagor (1993) that help motivate at-risk students; (b) as typical of most remedial students' papers, the editing exercises contained a variety of errors rather than focusing on a single error (Shaughnessy, 1977); (c) the cooperative editing exercises required collaboration (Johnson, Johnson, & Smith, 1991; Slavin, 1980). Shaughnessy claimed that when grammar lessons end with workbook exercises concentrating on the "feature being studied," they prevent the student from having "any practice in seeing that feature in more natural places" (p. 155). Indeed, such hands-on application in "natural places" better prepared the remedial students in the study to find and correct their own errors than lecture and textbook exercises did.

The significantly greater reduction in major grammatical errors by the cooperative editing students than by

the traditional students verifies the appropriateness of cooperative learning strategies for low-achieving students (Cuban, 1989; Davidson & Koppenhaver, 1993; Sagor, 1996; Slavin & DeVries, 1979). It cannot be concluded, however, that it was cooperative learning alone that accounted for the results of the study since cooperative learning was used with both experimental and control students via reader response groups. Rather, it was a particular type of cooperative learning, cooperative editing, that contributed most significantly to the results of the study. Although reader response groups did effectively generate encouragement and feedback about paper content from group members, they did not address grammatical concerns. Response groups also did not combine problem solving with group reward, as did cooperative editing.

The superior academic achievement of the cooperative editing students verifies the importance of integrating grammar with writing (Hairston, 1981; McCleary, 1995; Noguchi, 1991; Weaver, 1996), especially in classes with nontraditional students (Claywell, 1995; Delpit, 1995; Olson, 1995; Shaughnessy, 1977). Emphasizing the final editing stage in the composing process, the cooperative editing exercises not only trained students to detect and correct errors, but to defend and understand the choices they made. According to Harris and Rowan (1989), understanding is maximized when student writing is used to study the editing procedure.

In contrast to the claims of some researchers (Yager et al., 1985), the statistically significant positive result for the cooperative editing group indicates that academically homogeneous groups (all remedial composition students) can benefit from cooperative learning.

Pedagogically, cooperative editing has great potential. A simple procedure that promises good results, it should be adaptable to other levels of composition besides remedial college courses and to other aspects of composition besides standard grammar. Using the same process of cooperative editing of anonymous student papers, teachers could gear the exercises to the needs of the particular students involved, emphasizing mechanics, diction, style, research techniques, whatever seemed appropriate. In terms of teaching, one advantage of the cooperative editing exercises is the ease of grading. Requiring less than one-third the time that individual exercises require to grade, the cooperative editing exercises can be returned promptly for immediate feedback and discussion. Grading ease also means that exercises can be assigned (and graded) more frequently than possible with individual work. Additionally, cooperative editing should also prove an effective pedagogy for teachers of English methods courses as they prepare preservice teachers to mark and grade student papers.

The results of the study should be viewed in the context of certain limitations. External validity was

jeopardized by the small sample size which included only participants from one university in one region of the country. Thus, future researchers may choose to extend the study to universities in other regions and to other levels of students (e.g., middle school, secondary). The effects of similar interventions with students with disabilities and with students for whom English is not their primary language should also be investigated. Because the duration of the study was only one semester, a longitudinal study investigating the long-term effects of cooperative editing is needed.

Another limitation of the study was that all classes were taught by one of the researchers. However, the fact that the same instructor taught all of the classes increased the control of the study, ensuring that differences in teacher personality, appearance, and competency in error detection did not influence the outcome. Although the threat of researcher bias was minimized by the use of an objective dependent variable and by the blind grading process, future researchers should consider using a variety of instructors who are unaware of experimental conditions.

Additionally, future researchers who replicate the study might well choose to assess the affective dimension of the intervention via an attitudinal survey. They might also choose to compare the effects of cooperative editing versus individual editing of student papers.

References

- Beachy, C. J. (1992). Enhancing writing through cooperative peer editing. In N. Davidson & T. Worsham (Eds.), *Enhancing thinking through cooperative learning* (pp. 209-220). New York: Teachers College Press.
- Bennett, N., Jordan, J., Long, G., & Wade, B. (1976). *Teaching styles and student progress*. Cambridge, MA: Harvard University Press.
- Borg, W. R., & Gall, M. D. (1989). *Educational research: An introduction*. (5th ed.). New York: Longman.
- Braddock, R. R., Lloyd-Jones, R., & Shoer, L. (1963). *Research in written composition*. Champaign, IL: National Council of Teachers of English.
- Bruffee, K. A. (1993). *Collaborative learning: Higher education, interdependence, and the authority of knowledge*. Baltimore: Johns Hopkins.
- Claywell, G. (1995). Reasserting grammar's position in the trivium in American college composition. In S. Hunter & R. Wallace (Eds.), *The place of grammar in writing instruction* (pp. 43-53). Portsmouth, NH: Boynton/Cook.
- Connors, R. H., & Lunsford, A. A. (1988). Frequency of formal errors in current college writing, or Ma and Pa

- Kettle do research. *College Composition and Communication*, 39, 395-409.
- Cuban, L. (1989). At-risk students: What teachers and principals can do. *Educational Leadership*, 46(5), 29-32.
- Davidson, J., & Koppenhaver, D. (1993). *Adolescent literacy: What works and why*. (3rd ed.). New York: Garland Publishing.
- Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom*. New York: The New Press.
- Donovan, J. M. (1990). Use cooperative learning to teach grammar. Resurrect the dragon grammaticus. *English Journal*, 79, 62-65.
- Elbow, P. (1973). *Writing without teachers*. London: Oxford University Press.
- Fowler, H. R., Aaron, J. E., & Tavers, J. K. (1992). *The Little Brown handbook*. (5th ed.). New York: Harper Collins.
- Gebhardt, R. (1980). Teamwork and feedback: Broadening the base of collaborative writing. *College English*, 42, 69-74.
- Hairston, M. (1981). Not all errors are created equal: Nonacademic readers in the professions respond to lapses in usage. *College English*, 43, 794-806.
- Harris, M., & Rowan, K. E. (1989). Explaining grammatical concepts. *Journal of Basic Writing*, 8(2), 21-41.
- Hart, R. L. (1991). An investigation of the effects of collaborative learning on the writing skills of composition II students at Gloucester County College. Doctoral practicum report, Nova University. (ERIC Document Reproduction Service No. ED 341 058)
- Hartwell, P. (1985). Grammar, grammars, and the teaching of grammar. *College English*, 47, 105-127.
- Herreman, D. (1988). None of us is as smart as all of us. In *Focus on collaborative learning: Classroom practices in teaching English*. Urbana, IL: National Council of Teachers of English. (ERIC Document Reproduction Service No. ED 297 338)
- Hillocks, G. (1986). *Research on written composition*. Urbana, IL: National Conference on Research in English.
- Holden, M. (1994). *Effectiveness of two approaches to teaching writing in improving students' knowledge of English grammar*. (ERIC Document Reproduction Service No. ED 297-338)
- Hunter, S., & Wallace, R. (1995). *The place of grammar in writing instruction: Past, present, future*. Portsmouth, NH: Boynton/Cook.
- Hurst, M. J. (1987, Spring). Teaching students to write more effective sentences. *Arizona English Bulletin*, 29(3). (ERIC Document Reproduction Service No. ED 346 082)
- Johnson, D. W., Johnson, R. T., & Holubec, E. J. (1993). *Cooperation in the classroom*. Edina, MN: International Book Company.
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (1991). *Cooperative learning: Increasing college faculty instructional productivity*. ASHE-ERIC Higher Education Report No. 4. Washington, D.C.: The George Washington School of Education and Human Development.
- Langan, J. (1991). *English skills with readings*. (2nd ed.). New York: McGraw-Hill.
- Meyer, C. F. (1986). Improving instruction in grammar. *Journal of Teaching Writing*, 5(1), 17-31.
- McCleary, B. (1995). Grammar making a comeback in composition teaching. *Composition Chronicle: Newsletter of Writing Teachers*, 8(6), 1-4.
- Neulieb, J., & Brosnahan, I. (1987). Teaching grammar to writers. *Journal of Basic Writing*, 6, 28-35.
- Ney, J. W. (1989). *Teacher-student cooperative learning in the freshman writing course*. Tempe, AZ: Temple University. (ERIC Document Reproduction Service No. ED 312 659)
- Noguchi, R. R. (1991). *Grammar and the teaching of writing: Limits and possibilities*. Urbana, IL: National Council of Teachers of English.
- Office of Institutional Research. (1998). Mississippi State, MS: Mississippi State University.
- Olson, J. (1995). A question of power: Why Frederick Douglass stole grammar. In S. Hunter & R. Wallace (Eds.), *The place of grammar in writing instruction* (pp. 30-42). Portsmouth, NH: Boynton/Cook.
- Presseisen, B. Z. (1992). The evolution of cooperative thinking. In N. Davidson & T. Worsham (Eds.), *Enhancing thinking through cooperative learning* (pp. 1-6). New York: Teachers College Press.
- Renwick, M. K. (1994). Real research into the real problems of grammar and usage instruction. *English Journal*, 83, 29-32.
- Rose, C. A. (1996). The great debate: Teaching grammar and usage. *English Journal*, 85, 96-97.
- Sagor, R. (1993). *At-risk students: Reaching and teaching*. Swampscott, MA: Western Publishing Co.
- Shaughnessy, M. P. (1977). *Errors and expectations: A guide for the teacher of basic writing*. New York: Oxford University Press.
- Slavin, R. E. (1980). Cooperative learning. *Review of Educational Research*, 50(2), 315-342.
- Slavin, R. E. (1987). Cooperative learning and the cooperative school. *Educational Leadership*, 45, 7-13.
- Slavin, R. E., & DeVries, D. (1979). Learning in teams. In H. J. Waiberg (Ed.), *Educational environments and effects* (pp. 120-139). Berkeley, CA: McCutchan.

- Talmage, H., Pascarella, E. T., & Ford, S. (1984). The influence of cooperative learning strategies on teacher practices, student perceptions of the learning environment, and academic achievement. *American Educational Research Journal*, 21, 163-179.
- Thompson, E. H. (1988). Ensuring the success of peer revision groups. *Focus on collaborative learning: Classroom practices in teaching English*. Urbana, IL: National Council of Teachers of English. (ERIC Document Reproduction Service No. ED 297 338)
- Vavra, E. (1985). *A new method and curriculum for teaching writing and grammar*. Paper presented at the National Conference of the Society of Educators and Scholars, Teaneck, NJ. (ERIC Document Reproduction Service No. ED 277 003)
- Weaver, C. (1996). *Teaching grammar in context*. Portsmouth, NH: Boynton/Cook.
- Yager, S., Johnson, D. W., & Johnson, R. T. (1985). Oral discussion, group-to-individual transfer, and achievement in cooperative learning groups. *Journal of Educational Psychology*, 77, 60-66.

APPENDIX A PROOFREADING

Verbs:

S-V agreement - Underline subjects once, verbs twice. Do they agree in number? Be sure that you use an *s* ending on present tense singular verbs (except with *I* or *you* as the subject).

Tense - Be sure that you have used the proper ending to indicate past tense (no helping verb) or the past participle (must have a helper, a form of *have* or *be*). Check for *ed* endings on these forms of regular verbs. Do not use the past participle with the following helping verbs: do, will, may, can, might, could, should, would. These helping verbs can be used with present tense forms of the verb. (*I did go*; *I might play*; *I should study*, etc.)

Sentence sense and major punctuation:

CS, FS - Identify independent and dependent clauses. Do you have a period or semicolon between 2 independent clauses unless they are joined by a coordinate conjunction (and, but, or, nor, for, so, and yet)?

Fragment - After identifying dependent and independent clauses, be sure that each group of words punctuated as a sentence has at least one independent clause. Do not punctuate any phrases or dependent clauses as sentences.

Pronouns (and possessive nouns):

Case - Check all personal pronouns and who and whom for proper case. Be sure that nominative case (*I, we, he, she, it, they, who*) is used for subjects and predicate nominatives, that objective case (*me, us, him, her, it, them, whom*) is used for all objects (direct object, indirect object, object of preposition) and that possessive case is used for any pronoun that shows possession (*me, our, his, her, its, their*). Also check possessive nouns for apostrophes.

P-A agreement - Circle each pronoun and draw an arrow to its antecedent. Do they agree in number? Remember that *each, either, neither, anyone, everyone*, and other indefinite pronouns are singular.

Pronoun reference - Do all pronouns, particularly *which, this, it, and they*, have specific, stated antecedents? Are any ambiguous?

APPENDIX B COOPERATIVE EDITING EXERCISE

Directions: Identify and correct the errors in the following paper; write the name of the error and its corresponding rule over each correction. Highlight corrections, names of errors, and the corresponding rules. Sign your names to the copy turned in by your group.

Why I Want to be an Inspiration to Children

Children are the most precious gift God gave to us. It does not matter what race, background, or heritage they come from, they are all special in my eyes.

When I started to high school, I made many goals for my future. One of them was to start a cheerleading school for the kids at the Girls Club. An after school care where girls whose parents both work can go after school. The Girls Club is very rewarding to me, because the girls there usually didn't have a lot of money. I was always blessed by having the opportunity to take dance, cheerleading, and softball. The girls there never had that opportunity, I decided to give them a chance.

I went to the club every Friday afternoon. We start off by playing games, I got to know the girls better by having fun with them first. The games were fun, but it was the cheerleading lessons they loved. It was such an inspiration to me to see their little faces glow after they done a new stunt.

The girls that I worked with were so very talented that I decided to enter them in the Christmas parade. We

worked extremely hard for a month preparing for the big event we got many compliments on the girls and how talented they were.

This year I am going to teach a dance class at a local studio. Dancing and cheerleading is what I do best, I hope I can bring as much love, joy, and happiness to my new students' lives as I done for the ones at the club.

APPENDIX C RULES FOR EDITING

Case - Violation of proper case form (nominative, objective, possessive) based on use of noun or pronoun in the sentence

- 6 general rule for use of subjective (nominative), objective, and possessive case, pages 212-214
- 6a compound subjects and complements (after linking verb), page 214
- 6b compound objects, page 215
- 6c *we* or *us* with nouns, page 216
- 6d pronouns as appositives, page 216
- 23 possessive nouns

Tense - Incorrect tense ending (failure to use *ed*, *en*, for example) or failure to use needed helping verb

- 7a use of principal parts of verbs, page 226
- 7b use of *lie*, *lay*, *sit*, and *set*, page 230
- 7c omitted *ed* endings, page 230
- 7d omitted helping verbs, page 231

Subject-Verb Agreement (S-V agr.) - Disagreement in the number (singular or plural) of subject and verb

- 8a-1 *s* ending on verbs with 3rd person singular subjects, page 247
- 8a-2 words in between subject and verb, page 249
- 8a-3 subjects joined by *and*, page 249
- 8a-4 subjects joined by *or* or *nor*, page 250
- 8a-5 indefinite pronouns as subjects, page 250
- 8a-6 collective (group) nouns as subjects, page 251
- 8a-7 subject comes *after* verb (usually *there* at beginning), page 252

8a-8 sentences with linking verbs, page 252

8a-9 *who*, *which*, and *that* used as subjects of dependent clauses, page 253

8a-10 nouns with plural forms but singular meaning, page 253

8a-11 titles or words named as words as subjects, page 254

Pronoun-Antecedent Agreement (P-A agr.) - A pronoun which does not agree in number with the word to which it refers (its antecedent)

- 8b-1 antecedents joined by *and*, page 256
- 8b-2 antecedents joined by *or* or *nor*, page 256
- 8b-3 indefinite pronouns as antecedents (*each*, *either*, *everyone*, *somebody*, etc.), page 256
- 8b-4 antecedents that are collective nouns (groups), page 258

Fragment (Frag.) - A phrase or dependent clause (incomplete thought) punctuated as a sentence

- 10 tests for fragments; revision, pages 272-275

Comma Splice (CS) - Two complete sentences joined together with a comma and no coordinate conjunction

- 11a main clauses with no conjunction joined by a comma, page 283
- 11b main clauses with no conjunction joined by a comma and a conjunctive adverb (transition), page 286

Fused Sentence (FS) - Two complete sentences with no mark of punctuation between them

- 11c two main clauses with no mark of punctuation between them, page 288

Pronoun Reference (Ref.) - A pronoun with no clear noun to which it refers

- 12a reference to one clear antecedent, page 291
- 12b reference to a close antecedent, page 293
- 12c reference to a specific rather than an implied noun, page 294

**APPENDIX D
ERROR CHARTS**

CLASS ERROR CHART

PAPER NUMBER _____ TYPE _____

NAME	CS	FS	FRAG	S-V	TENSE	P-A	CASE	REF	GRADE
TOTALS									

STUDENT ERROR CHART

PAPER	CS	FS	FRAG	S-V	TENSE	P-A	CASE	REF	GRADE
ONE									
TWO									
THREE									
FOUR									
FIVE									
SIX									
SEVEN									
EIGHT									
TOTALS									

Factors Associated with Achievement in Educational Research Courses

Anthony J. Onwuegbuzie
Valdosta State University

John R. Slate
University of Texas at El Paso

Frances R. A. Paterson, and Mary H. Watson
Valdosta State University

Robert A. Schwartz
Florida State University

The purpose of this study was to investigate empirically factors related to achievement in educational research courses. Participants were 121 graduate students enrolled in an educational research course at a southeastern university. Multiple regression analyses revealed that students with the lowest levels of achievement tended to be or to exhibit at least one of the following characteristics: younger, lower in overall academic achievement, higher levels of research anxiety, higher levels of statistics anxiety associated with computation self-concept, poorer study habits, lower expectations for their overall achievement in the research course, more previous research methodology courses, and heaviest course loads. These variables explained 50.5% of the variance in achievement, indicating a large effect size. Implications of these findings are discussed, as are suggestions for future research.

The graduate-level educational research course is particularly important for non-doctoral students, because it often represents their only formal exposure to research methodologies. Their experiences in this course may cement their attitudes towards the field of research, and thus determine whether they become consumers of research in the future (Onwuegbuzie, DaRos, & Ryan, 1997; Ravid & Leon, 1995). However, although the goal of many educational research instructors is to produce students with

research consumer skills (i.e., the ability to read, to interpret, to synthesize, and to utilize research) and research production skills (i.e., the ability to design and to implement original research studies) (Ravid & Leon, 1995), many students who have completed research methodology courses report being inadequately prepared either to understand or to conduct research (Fleming, 1988; Green & Kvidahl, 1990; Rackliffe, 1988). Moreover, many students typically experience lower levels of performance in these courses than in their other graduate-level classes. These students often view educational research courses as merely a hurdle which they must overcome in order to obtain their degrees (Onwuegbuzie, in press).

Surprisingly, only a few studies have been conducted in which correlates of achievement in educational research courses have been investigated. Recently, however, Onwuegbuzie and Daley have investigated the role of learning styles and anxiety in educational research courses (Onwuegbuzie, 1997a; Onwuegbuzie & Daley, 1997, 1988a). In particular, Onwuegbuzie and Daley (1997) found that graduate students ($n = 74$) who are able to evaluate research articles successfully maintain positive attitudes towards the presence of authority figures (semi-partial $r^2 = 26.7\%$). Also, students adept at proposal writing prefer working alone in quiet environments (semi-partial $r^2 = 24.9\%$). Finally, students who demonstrate the most knowledge of research concepts, methodologies, and

Anthony J. Onwuegbuzie is an Assistant Professor in the Department of Educational Leadership at Valdosta State University, Georgia. John R. Slate is a Professor in the Department of Educational Leadership at University of Texas at El Paso. Frances R. A. Paterson is an Assistant Professor in the Department of Educational Leadership at Valdosta State University, Georgia. Mary H. Watson is director of Grants and Contracts at Valdosta State University and an Assistant Professor in the Department of Educational Leadership. Robert A. Schwartz is an Associate Professor in the Department of Educational Leadership at Florida State University, Florida. The authors would like to express their gratitude to Christine E. Daley for her editorial assistance and to the four anonymous reviewers for their extremely helpful comments on an earlier version of this article. Correspondence should be addressed to Anthony Onwuegbuzie, Department of Educational Leadership, College of Education, Valdosta State University, Valdosta, Georgia, 31698 or by e-mail to <TONWUEGB@VALDOSTA.EDU>.

applications are those who prefer less course structure (semi-partial $r^2 = 5.9\%$), and who prefer working alone (semi-partial $r^2 = 27.4\%$), learning via the use of multiple resources (semi-partial $r^2 = 5.5\%$), and working in the evening (semi-partial $r^2 = 6.4\%$). In addition, these students tend to require caloric intake during study (semi-partial $r^2 = 5.3\%$) and are not kinesthetically-oriented (semi-partial $r^2 = 4.9\%$). In a follow-up investigation, these authors reported that graduate students ($n = 137$) who are the most similar in learning style to their instructor with respect to persistence orientation, peer orientation, auditory preference, and multiple perceptual preferences tend to obtain higher scores on the following three achievement measures: (1) evaluating research articles, (2) writing research proposals, and (3) conceptual knowledge (Onwuegbuzie & Daley, 1998a). The effect size pertaining to this relationship, as measured by the multi-variate canonical correlation, was .52. In addition, combining quantitative and qualitative techniques, Onwuegbuzie (1997a) concluded that anxiety debilitates students' ability to write research proposals adequately.

Although many students continue to struggle in educational research courses, no study appears to have been conducted that examines simultaneously several potential predictors of research achievement. The present investigation was designed for this purpose. That is, we examined the role of various cognitive, affective, personality, and demographic variables in predicting achievement in these courses. The following cognitive variables were used: overall academic achievement (as measured by grade point average), study habits, and students' expectations of their performance in the educational research course (measured on a 100-point scale). The affective variables comprised research anxiety, statistics anxiety, and composition anxiety. Finally, the following demographic factors were considered: age, number of college-level research methodology courses taken, number of college-level statistics courses taken, number of college-level mathematics courses taken, time elapsed since previous college-level mathematics course, course load, and students' occupational status.

Literature Review

Cognitive Variables

According to Lindner and Harris (1992a), self-regulated learning is defined as the integration and utilization of cognitive, metacognitive, motivational, perceptual, and environmental components in the successful completion of academic tasks. Self-regulated learning appears to be an important aspect of college academic achievement. Indeed, a statistically significant relationship between self-regulated learning and overall academic achievement consistently has been reported in the literature at both the

college (Lindner & Harris, 1992a, 1992b, 1993a, 1993b; Lindner, Harris, & Gordon, 1996; Paterson, 1996) and high school (Williams, 1996a, 1996b) levels. For example, Lindner and Harris (1992a) reported Pearson correlation coefficients which ranged from .31 to .54 ($n = 160$). In a two-part investigation, Lindner et al. (1996) reported correlation coefficients ranging from .18 to .30 for undergraduate students in their first study ($n = 96$), and from .17 to .44 for undergraduate students in their second study ($n = 281$).

Lindner and Harris (1992b) theorized that students become increasingly self-regulated as learners over the course of their college lives. Lindner et al. (1996) found that graduate students tend to report higher self-regulating scores than do undergraduates. Moreover, Lan (1996) reported that graduate students in statistics classes who were assigned to a self-monitoring group performed better on course examinations, utilized more self-regulated learning strategies, and demonstrated a more in-depth knowledge of the course content than did students assigned to either instructor-monitoring or control groups. These findings, taken together with the result that self-regulated learners also appear to be more adept at writing research proposals (Onwuegbuzie, 1997a), suggest that the same students who typically experience high levels of overall academic success also tend to attain high performance levels in educational research courses. Thus, we hypothesized that overall academic achievement, as measured by grade point average, would be related positively to achievement in educational research courses. In fact, college grade point average has been found to be related positively to performance in statistics classes (Ware & Chastain, 1989), as well as in other courses (Dambrot, Watkins-Malek, Silling, Marshall, & Garver, 1985).

According to Jones, Slate, and Kyle (1992), college students with high levels of overall academic achievement tend to have better study habits than do low-achieving students with respect to time management, study techniques, and attitudes towards learning. Indeed, a consistent positive relationship between study skills and academic performance has been reported in the literature (Al-Hilawani & Sartawi, 1997; Blustein et al., 1986; Jones & Slate, 1992). Jones, Slate, Perez, and Marini (1996), based on a series of studies conducted by Jones and Slate (1992), concluded that study skills account for approximately 15% of the variance in undergraduate students' grades. Moreover, Polansky, Horan, and Hanish (1993) found that training in study skills significantly increased the retention rate of at-risk college students. These findings, as well as others, highlight the importance of study skills with respect to academic achievement for college students. Thus, assuming that students with relatively low levels of performance in educational research courses tend to be those who attain lower levels of

performance in their other graduate courses than do their counterparts, we hypothesized that there would be a positive relationship between study skills and research methodology achievement.

Many students enroll in educational research classes expecting to fail or at least to underachieve due to negative self-perceptions about their ability (Onwuegbuzie, 1997a). Apparently, these students continually engage in negative, self-defeating ruminations, which make it difficult for them to focus on the task at hand. In addition, these students tend to utilize less efficient problem-solving strategies, which, in turn, threaten their performance levels in educational research courses (Onwuegbuzie, 1997a). In this respect, a self-fulfilling prophecy appears to operate. The converse also appears to be true. That is, students who expect relatively high levels of performance in educational research courses may tend to behave in ways that promote success (e.g., utilize effective learning strategies and study habits). Thus, we hypothesized that students' expectations of their achievement in educational research courses would be positively related to their ultimate performance. Indeed, such a relationship has been found in statistics courses ($r = .33$, $n = 278$; Feinberg & Halperin, 1978) and foreign language classes (Pearson $r = .35$, semi-partial $r^2 = 5.0\%$, $n = 184$; Onwuegbuzie, Bailey, & Daley, 1998).

Affective Variables

Students taking their initial graduate-level educational research course typically have had little or no formal exposure to statistics (Onwuegbuzie, 1997a). As such, these students often report high levels of statistics anxiety. In fact, this type of anxiety appears to be a barrier to success, not only in statistics courses (Benson, 1989; Onwuegbuzie & Daley, 1996; Onwuegbuzie & Seaman, 1995; Zeidner, 1991), but also in research methodology courses (Onwuegbuzie, 1997a). According to Onwuegbuzie (1997a), statistics anxiety primarily affects a student's ability to understand research articles fully, as well as to analyze and to interpret statistical data. Students also report high levels of research anxiety and composition anxiety in educational research courses (Onwuegbuzie, 1997a). Research anxiety tends to make it difficult for students to understand and to apply the scientific method when proposing or conducting research, whereas composition anxiety typically impedes performance when students are engaged in writing research proposals or in documenting their research findings.

Because statistics anxiety, research anxiety, and composition anxiety appear to have strong negative statistical relationships to achievement in educational research courses, it is likely that any model that seeks to predict research methodology achievement and does not

include measures of situation-specific anxiety would be under-specified.

Demographic variables

Age has been shown to be related to achievement in statistics courses. Using an experimental design, Onwuegbuzie and Daley (1998b) reported that older students tend to experience lower levels of performance when taking statistics examinations under timed conditions than when time limits are not imposed. In addition, Sagaria (1989) observed that non-traditional students tend to obtain one letter grade higher than do traditional students in introductory statistics classes. Because statistics is an important component of most educational research courses, it is likely that age also is a factor in achievement in research methodology courses. On the one hand, older students may have an advantage over their younger counterparts in educational research courses, due to a stronger commitment to their graduate degree programs (resulting from a calculated decision to pursue their degrees after establishing themselves in a career). On the other hand, younger students may be more receptive to utilizing computers for conducting literature searches and analyzing statistical data, which, in turn, may enhance their performance levels relative to older students. Thus, we hypothesized that there would be a relationship between age and achievement in educational research courses, although we did not predict whether this relationship would be positive or negative.

Because statistics anxiety has been found to be related to the number of college-level mathematics courses completed and the time elapsed since students' last mathematics course (Benson, 1989; Tomazic & Katz, 1988; Wilson, 1997), and because statistics anxiety debilitates performance in research methodology courses (Onwuegbuzie, 1997a), we hypothesized that the number of college-level mathematics courses completed and time elapsed since students' last mathematics course also would be related to research methodology achievement. In addition, we predicted that the number of college-level statistics courses and number of prior research courses also would be related to performance in educational research classes.

Educational research courses often are rated by students as being among the most difficult in their degree programs (Onwuegbuzie, 1997a). Thus, it is likely that those who carry a heavy course load while enrolled in these classes will have insufficient time to devote to mastering research concepts. In fact, it is logical to expect that the more courses students are taking overall, the more likely they will be to experience lower levels of performance in educational research courses, due to competing demands for study time. Thus, course load was

included in the study as a potential predictor of research methodology achievement. Similarly, occupational status was included as a possible correlate.

Methods

Participants

The sample comprised 121 graduate students from a number of education disciplines (e.g., early childhood education, elementary education, middle grades, secondary education, speech language pathology, and psychology) who had enrolled in three sections of an introductory-level educational research course in the Fall 1997 quarter at a southeastern university. Participation was voluntary. In order to participate, students were required to give their consent by signing an informed consent document. Subjects received extra course credit. All surveys were coded using student identification numbers in order to maximize confidentiality.

The ages of the participants ranged from 21 to 51 (mean = 25.8, $SD = 6.9$), with a mean grade point average (based on graduate courses) of 3.58 ($SD = 0.36$). The majority of the sample was female (92.6%). With respect to ethnicity, the group comprised Caucasian-American (93.3%) and African-American (6.7%) students. Respondents had various degrees of exposure to mathematics, statistics, and research courses. Indeed, the number of college-level mathematics courses they had taken previously ranged from 0 to 18 ($M = 3.2$, $SD = 3.4$), the number of college-level research courses they had taken previously ranged from 0 to 6 ($M = 0.8$, $SD = 1.2$), and the number of college-level statistics courses they had taken previously ranged from 0 to 5 ($M = 0.6$, $SD = 1.0$). Approximately one-half of the sample (52.6%) had never taken a college-level research course, while the majority of the sample (85.9%) had never taken a college-level statistics course.

According to the university graduate handbook, the course "provides a study of methods used to conduct educational research and includes critical analysis of selected research articles and research design." Over the 10-week quarter, classes were held for four hours and ten minutes, once per week. The students' major course requirements involved (1) a written critical evaluation of a published research report (article critique) and (2) preparation of a research proposal. The major goal of the article critique was to allow students to practice evaluating published research articles utilizing principles of the scientific method. With respect to research proposals, the goal was to allow students the opportunity to practice the decision-making skills required for planning an original research study. Research proposals had to be unique, realistic, have educational significance, and extend the knowledge base. Whereas the article critiques were undertaken on an

individual basis, the research proposals were completed using cooperative learning groups. A scoring rubric was utilized to provide students with detailed feedback from their instructors for both assignments.

During the course, students were given three quizzes and midterm and final examinations. All examination questions attempted to assess higher-order conceptual processes. All test forms combined both open- and closed-response format items. Instruction was delivered in these courses using team-teaching techniques, with each section comprising two instructors. These instructors extensively used the available hardware (e.g., flexcams) and software (e.g., *Microsoft Powerpoint*) to present lecture material. Lecture was interspersed with group work activities, such that the course combined both didactic and constructivist approaches.

Instruments and Procedure

The following instruments were administered on the first day of class: the Statistical Anxiety Rating Scale, the Writing Apprehension Test, the Research Anxiety Rating Scale, the Study Habits Inventory, and the Background Demographic Form.

The Statistical Anxiety Rating Scale was developed by Cruise and Wilkins (1980). The Statistical Anxiety Rating Scale is a 51-item, 5-point Likert-format instrument which assesses statistics anxiety in a wide variety of academic situations. The Statistical Anxiety Rating Scale has six subscales, namely, *worth of statistics*, *interpretation anxiety*, *test and class anxiety*, *computation self-concept*, *fear of asking for help*, and *fear of the statistics instructor*. According to its authors, *worth of statistics* refers to students' perceptions of the relevance of statistics. *Interpretation anxiety* concerns the anxiety experienced when students are faced with making a decision from or interpreting statistical data. *Test and class anxiety* pertains to the anxiety involved when taking a statistics class or test. *Computation self-concept* involves the anxiety experienced when attempting to solve mathematical problems, as well as the students' perceptions of their ability to understand mathematics. *Fear of asking for help* measures the anxiety experienced when asking a fellow student or professor for help in understanding the material covered in class or any type of statistical data, such as an article or a printout. Finally, *fear of statistics teachers* is concerned with the student's perception of the statistics instructor. A high score on any subscale represents high anxiety in this area. Normative data have been gathered for this instrument. Cruise, Cash, and Bolton (1985) reported evidence of construct validity based on a factor analysis using 1,150 subjects in which six specific factors were identified after a varimax rotation. Loadings for these factors ranged from .48 to .86. These authors reported scores for each factor which had classical theory

alpha reliability coefficients ranging from .68 to .94 (median = .88). In addition, Cruise et al., using a sample of 161 students, reported five-week test-retest reliability coefficients for each factor, which ranged from .67 to .83 (median = .76). For the present study, the scores pertaining to each of the Statistical Anxiety Rating Scale subscales had classical theory alpha reliability coefficients as follows: *worth of statistics*, .96; *interpretation anxiety*, .90; *test and class anxiety*, .88; *computation self-concept*, .86; *fear of asking for help*, .81; and *fear of the statistics instructor*, .83.

The Writing Apprehension Test, developed by Daly and Miller (1975), is a 26-item, 5-point Likert-format instrument that assesses the degree to which individuals associate writing with apprehension, anxiety, and failure. Scores for the total scale range from 26 to 130, with high scores representing high levels of writing anxiety. Daly and Miller reported scores which had a classical theory Cronbach's alpha reliability of .94 ($n = 164$ undergraduate students enrolled in basic composition and interpersonal courses). Onwuegbuzie (1998b) reported scores which had an alpha reliability of .96 ($n = 89$ graduate students). For the current study, scores pertaining to the Writing Apprehension Test had a classical theory alpha reliability coefficient of .96.

The Research Anxiety Rating Scale, which was developed by Onwuegbuzie (1996), is a 45-item, 5-point Likert format instrument which assesses anxiety experienced by students in research methodology courses. Scores for the total scale range from 45 to 225, with high scores on the scale representing high levels of research anxiety. Using classical theory, Onwuegbuzie (1996) reported scores for the total scale which had a coefficient alpha reliability of .84 ($n = 262$ graduate students). For the present study, scores pertaining to the Research Anxiety Rating Scale had a classical theory alpha reliability coefficient of .94.

The Study Habits Inventory, developed by Jones and Slate (1992), consists of 63 true-false items designed to assess the typical study behaviors of college students. Thirty items describe effective study behaviors, and 33 items specify ineffective study behaviors. The latter items are key-reversed such that total scale scores range from 0 to 63, with high scores indicating good study skills. A series of studies conducted by Slate, Jones, and their colleagues suggests that the Study Habits Inventory yields reliable scores. In particular, using 128 graduate students, Jones et al. (1996) reported scores which generated an alpha reliability of .85; Slate, Jones, and Harlan (1998) reported scores which had an alpha reliability of .86 for

132 students at a post-secondary institute; Lawler-Prince, Slate, and Jones (1993) reported scores which yielded an alpha reliability of .84 for 146 college seniors majoring in elementary and/or special education at a university in the mid-south; and Jones et al. (1992) reported scores which had an alpha reliability of .84 for 115 secondary education preservice students at a mid-southern university. Validity of scores generated from the Study Habits Inventory has been established through significant correlations with college students' grades (Jones & Slate, 1992). For the current study, scores pertaining to the Study Habits Inventory yielded a classical theory alpha reliability coefficient of .85.

The Background Demographic Form, developed specifically for this study, extracted relevant demographic information such as age, sex, ethnicity, degree program, grade point average, number of college-level research methodology courses taken, number of college-level statistics courses taken, number of college-level mathematics courses taken, time elapsed since students' previous college-level mathematics course, course load, and students' occupational status.

Finally, achievement in the educational research course was measured using students' course averages. This global measure was selected rather than isolated measures of specific skills in order to maximize the external validity (i.e., generalizability) of the findings. Since students in all sections were presented with the same lecture material, were assigned identical projects graded with the same scoring rubrics, and were administered identical quizzes and examinations, students' scores within each section were not standardized. Students' course averages comprised evaluating research articles, writing research proposals, orally presenting research proposals, and conceptual knowledge (as measured by five untimed in-class examinations).

Results

Table 1 presents the correlations between each of the selected independent variables and overall educational research achievement, as well as the means and standard deviations of all variables. Using the Bonferroni adjustment (Maxwell & Delaney, 1990), it can be seen that achievement in educational research correlated negatively with the following variables: worth of statistics, test and class anxiety, computation self-concept, and course load. In addition, achievement correlated positively with study habits, age, and expected overall average for the current educational research course. Table 2 presents the inter-correlations between all of the predictor variables.

Table 1
Means and Standard Deviations of all Variables and Pearson
Product-Moment Correlations Between Cognitive, Affective, and
Demographic Variables and Educational Research Achievement

Variable	Mean	SD	Correlations With Educational Research Achievement
Outcome:			
Educational Research Achievement	88.51	4.92	—
Cognitive:			
Academic Achievement	3.58	0.36	.27
Study Habits	58.02	13.60	.35*
Expectation of Ed. Research Achievement	90.09	4.47	.47*
Affective:			
Research Anxiety	126.94	25.28	-.21
Composition Anxiety	73.69	19.93	-.10
Worth of Statistics	42.96	14.19	-.37*
Interpretation Anxiety	32.20	8.59	-.25
Test and Class Anxiety	27.16	7.31	-.30*
Computation Self-concept	18.47	6.25	-.41*
Fear of Asking for Help	8.30	3.68	-.13
Fear of the Statistics Instructor	12.27	4.04	-.19
Demographic:			
Age	25.81	6.86	.32*
Number of College-level Research Courses	0.79	1.20	-.23
Number of College-level Statistics Courses	0.58	1.04	.17
Number of College-level Mathematics Courses	3.22	3.50	.27
Time Elapsed since Previous College-level Math Course	59.68	49.30	.20
Course Load	3.23	0.44	-.32*
Students' Occupational Status	0.27	0.44	.29

* $p < .0028$ (i.e., after Bonferroni adjustment)

Table 2
Intercorrelations Among All Predictor Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Academic Achievement																	
2. Study Habits	.09																
3. Expectation of Achievement	.09	.22															
4. Research Anxiety	-.09	.52	-.28														
5. Composition Anxiety	-.18	-.39	-.13	.62													
6. Worth of Statistics	-.09	-.52	-.22	.69	.36												
7. Interpretation Anxiety	-.10	-.31	-.22	.63	.45	.53											
8. Test and Class Anxiety	-.12	-.33	-.20	.47	.17	.58	.66										
9. Computation Self-concept	-.13	-.40	-.22	.53	.25	.76	.58	.64									
10. Fear of Asking for Help	.08	-.39	-.16	.49	.24	.44	.57	.53	.41								
11. Fear of the Stats. Instructor	-.04	-.44	-.10	.63	.32	.75	.52	.54	.63	.50							
12. Age	.21	.29	.15	-.17	-.15	-.17	-.03	.02	-.05	-.03	-.18						
13. Number of Research Courses	-.13	.08	-.07	-.09	.08	.07	-.08	.04	.03	-.12	-.01	-.01					
14. Number of Statistics Courses	-.07	.11	.10	-.01	.15	-.17	-.10	-.28	-.22	-.01	-.07	.04	.29				
15. Number of Mathematics Courses	.05	.20	.13	-.26	-.14	-.31	-.36	-.35	-.39	-.25	-.29	.13	.29	.43			
16. Time Elapsed Since Math Course	.14	.24	.06	-.02	-.07	-.01	.03	.09	.11	.01	-.06	.61	-.01	.07	.00		
17. Course Load	-.08	-.19	-.04	.21	.22	.21	.14	.12	.21	.02	.02	-.24	-.01	-.09	-.25	-.12	
18. Students' Occupational Status	.22	.09	-.04	-.20	-.20	-.28	-.15	-.13	-.19	.01	.01	.36	-.13	.06	.39	.24	-.45

Note: Decimals omitted

The Shapiro-Wilk test (Shapiro & Wilk, 1965; Shapiro, Wilk, & Chen, 1968) did not indicate that the distribution of educational research achievement scores was non-normal ($W = .97, p > .05$), thereby justifying the use of multiple regression. In addition, evaluation of assumptions of linearity and homogeneity revealed no threat to multiple regression analysis.

All possible subsets (APS) multiple regression (Thompson 1995) was used to identify an optimal combination of cognitive, affective, and demographic variables (i.e., independent variables) that predicted achievement in the educational research course. Using this technique, all possible models involving some or all of the independent variables were examined. This method of analysis has been recommended by many statisticians (e.g., Thompson, 1995). Indeed, in APS regression, separate regressions are computed for all independent variables singly, all possible pairs of independent variables, all possible trios of independent variables, and so forth, until the best subset of independent variables is identified according to some criterion. For this study, the criterion used was the maximum proportion of variance explained (R^2), which provides an important measure of effect size (Cohen, 1988). An additional index used was Mallow's C_p (Myers, 1986; Sen & Srivastava, 1990).

Squared semi-partial correlation coefficients, also known as part correlations, represent the amount by which R^2 is reduced if a particular independent variable is removed from the regression equation. That is, squared semi-partial correlation coefficients express the unique contribution of the independent variable as a proportion of the total variance of the dependent variable (Cohen, 1988). Similarly, squared partial correlation coefficients represent the unique contribution of the independent variable as a proportion of R^2 . In this study, squared partial correlation coefficients, like R^2 , were used directly as effect size estimates, as recommended by Cohen (1988). According to Cohen, for multiple regression models in the behavioral sciences, squared partial correlation values between 2% and 12.99% suggest small effect sizes, values between 13% and 25.99% indicate medium effect sizes, and values of 26% and greater suggest large effect sizes. These same criteria were used to assess whether the proportion of variance explained by the independent variables, R^2 , was suggestive of a small, medium, or large effect.

Table 3 presents the unstandardized regression coefficients and intercept, the standard error of the unstandardized coefficients, the standardized regression coefficients, the structure coefficients, the squared semi-

partial correlations, the squared partial correlation coefficients, and the squared multiple correlation coefficient (R^2) of the chosen model, as well as tolerance statistics, variance inflation factors, and condition numbers. The APS multiple regression analysis revealed that a model containing eight variables provided the best fit to these data. In fact, the best nine-variable model, in which the number of statistics courses taken was added to the model, only increased the proportion of variance explained by 1.6%. In addition, Mallow's C_p was closer in value to the number of regressor variables (Myers, 1986; Sen & Srivastava, 1990) with the eight-variable solution than with any nine-variable solution.

The selected model indicated that the following eight variables contributed significantly ($F[8, 112] = 14.26, p < .0001$) to the prediction of educational research achievement: research anxiety, computation self-concept, study habits, age, course load, number of college-level research courses taken, expectation of educational research achievement, and grade point average (Table 3). These eight variables combined to explain 50.5% of the variation in educational research achievement. Using Cohen's (1988) criteria for assessing the predictive power of a set of independent variables in a multiple regression model, the proportion of variance explained indicates a large effect size, because it well exceeds 26%.

An inspection of the studentized residuals generated from the model (Myers, 1986) suggested that the assumptions of normality, linearity, and homoscedasticity were met. Using the Bonferroni adjustment, none of the studentized residuals suggested that outliers were present. Additionally, an examination of the tolerance statistics, the variance inflation factors, and the condition numbers of the selected regression model indicated strongly that no multicollinearity was present. Specifically, all variance inflation factors (Table 3), which indicate the extent to which the variance of an individual regression coefficient has been inflated by the presence of collinearity (Sen & Srivastava, 1990), are much less than 10, which is Myer's (1986) criteria for suspecting the presence of multicollinearity. Indeed, all the variance inflation factors were relatively close to unity, which indicates no relationship among the regressor variables. Condition numbers represent the ratio of the largest to the smallest eigenvalues, which, in turn, are measures of the strength of linear dependency among regressor variables. From Table 3, it can be seen that all condition numbers are much less than Myer's (1986) cut-off score of 1000, again suggesting that multicollinearity is not a feature of these data.

Table 3
Selected Multiple Regression Model for Predicting Statistics Achievement

Variable	Regression Coefficient	Standard Error	t-value	Standardized Regression Coefficient	Structure Coefficient	Squared Semi-Partial Correlation Coefficients (%)	Squared Partial Correlation Coefficients (%)	Tolerance Statistics	Variance Inflation Factors	Condition Numbers
INTERCEPT	78.98	15.77	5.01***	—	—	—	—	—	—	1.00
Research Anxiety	2.60	1.39	1.88	0.16	-0.30	4.6	8.6	0.58	1.71	3.47
Computation Self-concept	-2.86	0.82	-3.49**	-0.29	-0.58	12.2	19.7	0.66	1.51	7.60
Study Habits	12.88	5.41	2.38*	0.20	0.49	4.9	8.9	0.66	1.52	8.61
Age	0.18	0.10	1.79	0.13	0.45	6.6	11.8	0.79	1.27	14.17
Course Load	-0.36	0.12	-2.87**	-0.21	-0.45	4.0	7.5	0.86	1.16	16.02
Number of College-level Research Courses	-1.31	0.51	-2.58**	-0.18	-0.33	4.8	8.8	0.95	1.05	23.38
Expectation of Educational Research Achievement	0.74	0.14	5.22***	0.37	0.67	12.2	19.8	0.89	1.12	37.93
Academic Achievement	2.87	1.76	1.63	0.11	0.39	1.2	2.3	0.91	1.10	92.85

Model $R^2 = .505$, $F(8, 112) = 14.26$

Adjusted $R^2 = .469$

* $p < .05$, ** $p < .01$, *** $p < .001$

As recommended by Thompson (1994), several empirical internal replicability analyses were conducted to evaluate further the adequacy of the selected regression model. Specifically, a jackknife method was used (Crask & Perreault, 1977). This involved conducting 121 separate regression analyses (each fitting the eight-variable model), whereby each analysis involved dropping the i th participant until every subject had been eliminated exactly once. That is, each of the resultant 121 regression models utilized 120 subjects (i.e., $n-1$ subjects, where n = the total sample size). The 121 adjusted and unadjusted R^2 values which were generated from these models were examined for stability. The summary statistics pertaining to this analysis are presented in Table 4. Assuming that the sample estimates of the multiple correlation coefficients are normally distributed (as suggested by the closeness of the mean and median values for both adjusted and unadjusted estimates), it can be seen that the 95% confidence interval about the parameter estimate lies between 50.4% and 50.6% for the unadjusted R^2 and between 46.8% and 47.0% for the adjusted R^2 . Encouragingly, these intervals are not only very narrow, but they contain the estimates calculated using the complete data (i.e., $R^2 = 50.5\%$, adjusted $R^2 = 46.9\%$; c.f., Table 3)—suggesting that neither the adjusted R^2 nor the unadjusted R^2 are impacted by variations in the sample.

Finally, the following additional influence diagnostics were examined: (1) the number of estimated standard errors (for each regression coefficient) that the coefficient changes if the i th observation were set aside (i.e., DFBETAS); (2) the number of estimated standard errors that the predicted value changes if the i th point is removed from the data set (i.e., DFFITS); and (3) the reduction in the estimated generalized variance of the coefficient over

what would have been produced without the i th data point (i.e., COVRATIO). Using criteria recommended in the literature (e.g., Myers, 1986; Sen & Srivastava, 1990), no subject generated DFBETAS, DFFITS, or COVRATIO values that were large enough to indicate that (s)he represented an outlying observation—again suggesting sample invariance.

Table 4
Descriptive Statistics and 95% Percent Confidence Intervals about the Jackknife Estimates of the Adjusted and Unadjusted Regression Effect Size (Using 121 Resamples Each of Sample Size 120)

Summary Statistic	Unadjusted R^2	Adjusted R^2
Mean	50.49	46.92
Median	50.48	46.91
Mode	50.47	46.90
Standard Deviation	0.55	0.58
95% Lower Bound	50.39	46.82
95% Upper Bound	50.58	47.02
Full Model Effect Size ($n = 121$)	50.50	46.90

The regression model suggests that students with the lowest levels of performance in educational research courses tended to have at least one of the following eight characteristics: younger, lower overall academic achievers, higher levels of research anxiety, higher levels of statistics anxiety associated with computation self-concept, poorer study habits, lower expectations for their overall achievement in the educational research course, more previous research methodology courses, and heaviest course loads.

From the squared semi-partial coefficients (Table 3), it can be seen that computation self-concept and students'

expectations of their achievement were the best predictors of overall achievement, each explaining 12.2% of the variance. The squared partial coefficients for these variables (i.e., 19.7% and 19.8%, respectively) suggest a moderate effect size. These variables were followed, respectively, by age, study habits, number of college-level research courses, research anxiety, course load, and grade point average. The predictive power of these latter variables represented small effect sizes. An examination of the structure coefficients (Table 3), using a cutoff correlation of 0.3 recommended by Lambert and Durand (1975) as an acceptable minimum loading value, suggests that all eight variables made important contributions to the model (even grade point average, which explained the smallest proportion of variance). The fact that both the standardized and structure coefficients pertaining to all variables were noteworthy indicates that none of these constructs acted as suppressor variables (Thompson, 1998; Thompson & Borello, 1985).

Discussion

The purpose of this study was to explore the contribution of various cognitive, affective, and demographic variables to the prediction of achievement in educational research courses. The major finding in this study is that each class of variable appears to play a role in these courses. That is, it appears that cognitive, affective, and demographic variables are related to research methodology achievement. Indeed, most of the hypotheses were supported via the multiple regression model. Specifically, all three hypotheses involving cognitive variables, two of the three hypotheses relating to affective variables, and three of the seven hypotheses pertaining to demographic variables were supported.

Although the gender composition of the present sample appears to reflect that of most educational research courses, the fact that subjects were predominantly female and White is a limitation of the present study. As such, it is not clear how generalizable the findings of the present study are, suggesting a need for replication using more males and non-white graduate students. Notwithstanding, although these findings do not establish causality, it is apparent that achievement in educational research courses is a complex phenomenon, involving a multitude of variables.

The finding that composition anxiety was not related to research methodology achievement ($r = -.10$, Table 1) may have stemmed from the fact that none of the quizzes or examinations involved extensive writing. Indeed, these forms of assessment comprised either multiple-choice items or open-ended items that required relatively short responses (i.e., one or two sentences). As such, it is

likely that poor writers were not unduly challenged in terms of constructing their responses, which may have attenuated the relationship between composition anxiety and achievement. Additionally, one of the two term projects (i.e., the research proposal) was undertaken using cooperative learning groups. It is likely that the most capable writers--that is, those with the lowest levels of composition anxiety--were largely responsible for the writing component of the proposal, or at least played the biggest part in making the final revisions to the proposal. Thus, future researchers should investigate the relationship between composition anxiety and achievement under conditions whereby success depends to a greater extent on each student's ability to write, as appears to be the case in more traditional educational research courses.

The fact that computation self-concept was one of the two best predictors of achievement is consistent with Onwuegbuzie (1997a), who found that statistics anxiety debilitates performance in research methodology courses. However, it was somewhat surprising that computation self-concept emerged as a correlate of achievement rather than other dimensions of statistics anxiety, particularly test and class anxiety, which has been found to be the greatest source of statistics anxiety (Onwuegbuzie, 1998a). This finding may stem from students' misperceptions regarding the nature of research. Indeed, many students mistakenly believe that research, and especially the field of statistics, primarily involves conducting mathematical hand computations and manipulating formulae (Onwuegbuzie et al., 1997). Students with such false perceptions tend to exhibit high levels of statistics anxiety, in which they become so preoccupied with failure that their ability to suppress interfering thoughts and to attend continuously to task-relevant responses is severely impeded (Hembree, 1988; Wine, 1980). As a result, they may be unable to optimize cognitive operations, such as short- and long-term memory retrieval, in evaluative situations, resulting in underachievement (Sarason, 1972).

Students' expectations of their performance in the course was the next best predictor of research methodology achievement (semi-partial $r^2 = 12.2\%$, Table 3). This finding suggests that a self-fulfilling prophecy may be taking place, inasmuch as students who expect to underachieve often develop learned helplessness and, thus, act in ways that prevent them from maximizing their performance levels (Onwuegbuzie, 1997a). One behavior that is indicative of learned helplessness is procrastination. Procrastinatory behavior may range from serious deliberations over undertaking tasks and assignments in educational research courses (Onwuegbuzie, 1997a) to delaying enrollment in educational research courses (Onwuegbuzie, 1997b).

Study skills also appears to have a moderate effect size (semi-partial $r^2 = 4.8\%$). The positive relationship between study skills and research methodology achievement is consistent with the relationship repeatedly found in the literature between study skills and overall academic achievement (Al-Hilawani & Sartawi, 1997; Blustein et al., 1986; Jones & Slate, 1992; Jones et al., 1992; Jones et al., 1996). Interestingly, the zero-order correlation between study skills and achievement in the current study (.35) falls in the middle of the range that typically is reported for correlations between study skills and overall academic achievement (Jones & Slate, 1992). Thus, future research should address whether attempts at improving students' study skills may help to increase performance levels in educational research courses.

An extremely surprising finding was the negative relationship found between the number of college-level research courses and achievement (semi-partial $r^2 = 4.8\%$). One would expect that previously having taken research courses would be a tremendous advantage for students. Yet, remarkably, it appears to have been a disadvantage. However, this apparent anomaly may have resulted from students who had taken more classes not putting forth the effort needed to achieve success in the class due to a false assumption that knowledge gained in previous research courses would guarantee success. Additionally, it is possible that what students deemed to be a research class represented a misnomer. For example, some research classes taught by instructors who are not research methodologists often involve little more than informal discussions of articles; yet, students may be given the false impression that these courses prepare them for graduate-level educational research courses. Indeed, many students are surprised by how much rigor is involved in formally critiquing research articles, acknowledging that "critiques" undertaken in their other classes were superficial in comparison. More information regarding the college-level courses termed "research courses" would have helped to explain further this unexpected finding.

The negative relationship found between research anxiety and achievement in educational research courses (semi-partial $r^2 = 4.6\%$) is consistent with Onwuegbuzie (1997a). These two findings combined suggest that high levels of research anxiety interfere with performance by impeding a student's ability to apply new knowledge to a specific problem, by making it difficult to recall what has been learned, and by hindering effective use of problem-solving strategies (Tobias, 1977, 1980).

An extremely interesting finding was the positive relationship found between age and achievement (semi-partial $r^2 = 6.6\%$). To the extent that this relationship is causal, it is possible that older students have a stronger commitment to their graduate degree programs, and hence to acquiring knowledge in educational research courses.

Such commitments promote effective self-regulatory learning behaviors, culminating in higher levels of performance. Also, it is possible that older students, in an attempt to compensate for any knowledge deficits arising from their hiatuses from formal education, work more diligently in the class, maximizing their success in educational research courses. Clearly, more research is needed in this area to determine the role of age in these classes.

The positive relationship found between course load and achievement (semi-partial $r^2 = 4.0\%$), although not necessarily a causal one, suggests the importance of dissuading students from taking too many classes concurrent with their educational research courses. Although it is beyond the scope of the present study to test, it is likely that the relationship between semester course load and achievement is causal, because students find educational research courses to be among the most difficult in their graduate programs (Onwuegbuzie, in press). Simply put, taking too many classes in conjunction with educational research likely limits how much time students are able to spend on the latter, and thus reduces performance levels. Researchers in the future should investigate whether the number of additional classes taken or the difficulty of these classes plays the biggest role in reducing performance in educational research courses.

Finally, the relationship between overall academic achievement (as measured by grade point average) and research methodology achievement (semi-partial $r^2 = 1.2\%$) suggests that overall academic competence is helpful in educational research courses. According to Jones et al. (1996), students with high grade point averages are more adept at allocating their time among various subjects. Apparently, this trait may help them to use their study skills more effectively than do their low-achieving counterparts. Furthermore, Jones et al. (1996) reported that students with high grade point averages are more likely to integrate their learning experiences; that is, using knowledge gained from other courses to assist them in understanding concepts in subsequent courses. Thus, it is likely that these attributes help to increase the performance levels of high academic achievers in educational research courses. However, the fact that this relationship was relatively small serves to highlight the uniqueness of educational research courses in graduate students' programs of study because it suggests that achievement in other courses is helpful, but not essential, for success in research methodology courses. Nevertheless, it should be noted that it is possible that the small effect size associated with this relationship stems from the restricted range of the grade point averages ($SD = .36$). That is, the effect size estimate pertaining to grade point average may represent an underestimate, since the sample was relatively homogeneous with respect to overall academic achievement. Simply put, a more heterogeneous sample may have led to a larger effect size. Thus, future research

should investigate further the role of academic achievement using more academically-heterogeneous samples.

In summary, achievement in educational research courses appears to be a function of cognitive, affective, and demographic variables. The variables selected in the regression model explained more than 50% of the variance in educational research achievement, suggesting a very large effect size. The internal replicability analyses undertaken (e.g., jackknife estimates, DFFITS, DFBETAS, COVRATIO) suggest that the proportion of variance explained by the model is somewhat sample invariant. However, it must be acknowledged that caution should be exercised in generalizing the findings of the present study due to the context-specificity of regression weights, as such, external replications are needed (Thompson, 1998).

Nevertheless, the present study represents an attempt to identify factors which may affect directly performance in these courses. A possible next step in the process should be to design some form of experimental study to determine the causal nature of these effects. Because the ultimate goal of educational research instructors is to maximize student understanding, and hence performance levels, interventions designed to target the "at-risk" students identified in this study should be developed and implemented differentially in accordance with students' individual diagnostic profiles. For example, students who exhibit high levels of research anxiety or statistics anxiety could be assigned to treatment groups employing anxiety-reduction procedures, such as mental and emotive imagery, relaxation therapy, systematic desensitization, cognitive and covert modeling, thought stopping, cognitive restructuring, meditation, biofeedback, and neuro-linguistic programming (Gilliland & James, 1983). Similarly, students with poor study habits could be given study skills training specifically tailored for research methodology classes. The effects of combining both anxiety-reducing techniques and study skills training could even be investigated, as recommended by Onwuegbuzie and Daley (1996), because this combination has been found to be more effective in raising performance levels than has using either method of treatment alone (Allen, 1971; Mitchell & Ng, 1972). In any case, assessing the effects of such interventions will yield useful information for instructors of educational research courses.

References

- Al-Hilawani, Y. A., & Sartawi, A. A. (1997). Study skills and habits of female university students. *College Student Journal*, 31, 537-544.
- Allen, G. J. (1971). Effectiveness of study counseling and desensitization in alleviating test anxiety in college students. *Journal of Abnormal Psychology*, 77, 282-289.
- Benson, J. (1989). Structural components of statistical test anxiety in adults: An exploratory model. *Journal of Experimental Education*, 57, 247-261.
- Blustein, D., Judd, T., Krom, J., Viniar, B., Padilla, E., Wedemeyer, R., & Williams, D. (1986). Identifying predictors of academic performance of community college students. *Journal of College Student Personnel*, 27, 242-249.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: John Wiley.
- Crask, M. R., & Perreault, W. D. (1977). Validation of discriminant analysis in marketing research. *Journal of Marketing Research*, 14, 60-68.
- Cruise, R. J., Cash, R. W., & Bolton, D. L. (1985, August). *Development and validation of an instrument to measure statistical anxiety*. Paper presented at the annual meeting of the Statistical Education Section. Proceedings of the American Statistical Association.
- Cruise, R. J., & Wilkins, E. M. (1980). *STARS: Statistical Anxiety Rating Scale*. Unpublished manuscript, Andrews University, Berrien Springs, MI.
- Daly, J. A., & Miller, M. D. (1975). The empirical development of an instrument to measure writing apprehension. *Research in the Teaching of English*, 9, 242-249.
- Dambrot, F. H., Watkins-Malek, M. A., Silling, S. M., Marshall, R. S., & Garver, J. A. (1985). Correlates of sex differences in attitudes toward and involvement with computers. *Journal of Vocational Behavior*, 27, 71-86.
- Feinberg, L. B., & Halperin, S. (1978). Affective and cognitive correlates of course performance in introductory statistics. *Journal of Experimental Education*, 46, 11-18.
- Fleming, D. (1988, April). *The literature on teacher utilization of research: Implications for the school reform movement*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Gilliland, B. E., & James, R. K. (1983). Hypnotherapy and cognition: A combinatorial approach. *Medical Hypnoanalysis: Journal of the Society of Medical Hypnoanalysts*, 4, 101-113.
- Green, K. E., & Kvidahl, R. F. (1990, April). *Research methods courses and post-bachelor's education: Effects on teachers' research use and opinions*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

- Hembree, R. (1988). Correlates, causes, effects and treatment of test anxiety. *Review of Educational Research*, 58, 47-77.
- Jones, C. H., & Slate, J. R. (1992). *Technical manual for the Study Habits Inventory*. Unpublished manuscript, Arkansas State University, Jonesboro, AR.
- Jones, C. H., Slate, J. R., & Kyle A. (1992). Study skills of teacher education students. *Teacher Education*, 28, 7-15.
- Jones, C. H., Slate, J. R., Perez, E., & Marini, I. (1996). Graduate students' study skills as a function of academic achievement, sex, conceptions of intelligence, and locus of control. *New Directions for Education Reform*, 3, 61-78.
- Lambert, Z. V., & Durand, R. M. (1975). Some precautions in using canonical analysis. *Journal of Market Research*, XII, 468-475.
- Lan, W. Y. (1996). The effects of self-monitoring on students' course performance, use of learning strategies, attitude, self-judgement ability, and knowledge representation. *Journal of Experimental Education*, 64, 101-115.
- Lawler-Prince, D., Slate, J. R., & Jones, C. H. (1993). Academic behaviors of preservice elementary and/or special education teachers: A preliminary study. *Louisiana Educational Research Journal*, 18, 109-118.
- Lindner, R. W., & Harris, B. (1992a, April). *Self-regulated learning and academic achievement in college students*. Paper presented at the American Educational Research Association annual meeting, San Francisco, CA. (ERIC Document Reproduction Service No. ED 345 626)
- Lindner, R. W., & Harris, B. (1992b). *The development and evaluation of a Self-Regulated Learning Inventory and its implication for instructor-independent instruction*. (ERIC Document Reproduction Service No. ED 348 010)
- Lindner, R. W., & Harris, B. (1993a, January). *Teaching self-regulated learning strategies*. Paper presented at the Convention of the Association for Educational Communications and Technology, New Orleans, LA. (ERIC Document Reproduction Service No. ED 362 182)
- Lindner, R. W., & Harris, B. (1993b). Self-regulated learning: Its assessment and instructional implications. *Educational Research Quarterly*, 16, 29-37.
- Lindner, R. W., Harris, B. R., & Gordon, W. I. (1996, April). *Are graduate students better self regulated learners than undergraduates? A follow-up study*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY. (ERIC Document Reproduction Service No. ED 396 638)
- Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.
- Mitchell, K. R., & Ng, K. T. (1972). Effects of group counseling and behavior therapy on the academic achievement of test-anxious students. *Journal of Counseling Psychology*, 19, 491-497.
- Myers, R. H. (1986). *Classical and modern regression with applications*. Boston, MA: Duxbury Press.
- Onwuegbuzie, A. J. (1996). *Development of the Research Anxiety Rating Scale*. Unpublished manuscript, University of Central Arkansas.
- Onwuegbuzie, A. J. (1997a). Writing a research proposal: The role of library anxiety, statistics anxiety, and composition anxiety. *Library & Information Science Research*, 19, 5-33.
- Onwuegbuzie, A. J. (1997b). The teacher as researcher: The relationship between enrollment time and achievement in a research methodology course. *Reflection and Research*, 3(1)[On-Line]. Available: <http://www.soc.gonzaga.edu/rr/v3n1/tony.html>
- Onwuegbuzie, A. J. (1998a). The dimensions of statistics anxiety: A comparison of prevalence rates among mid-southern university students. *Louisiana Educational Research Journal*, 23, 23-40.
- Onwuegbuzie, A. J. (1998b). The relationship between writing anxiety and learning styles among graduate students. *Journal of College Student Development*, 39, 589-598.
- Onwuegbuzie, A. J. (in press). Teachers' attitudes towards educational research courses: Implications for the teacher-as-researcher movement. *GATEways to Teacher Education*.
- Onwuegbuzie, A. J., Bailey, P., & Daley, C. E. (1998). *Cognitive, affective, personality, and demographic predictors of foreign language achievement*. Article submitted for publication.
- Onwuegbuzie, A. J., & Daley, C. E. (1996). The relative contributions of examination-taking coping strategies and study coping strategies on test anxiety: A concurrent analysis. *Cognitive Therapy and Research*, 20, 287-303.
- Onwuegbuzie, A. J., & Daley, C. E. (1997). Learning style and achievement in a course on research methods. *Psychological Reports*, 80, 496-498.
- Onwuegbuzie, A. J., & Daley, C. E. (1998a). Similarity of learning styles of students and a teacher in achievement in a research methods course. *Psychological Reports*, 82, 163-168.
- Onwuegbuzie, A. J., & Daley, C. E. (1998b). Age-related differences in timed examinations. *Louisiana Educational Research Journal*, 23, 23-40.
- Onwuegbuzie, A. J., DaRos, D., & Ryan, J. (1997). The components of statistics anxiety: A phenomenological

- study. *Focus on Learning Problems in Mathematics*, 19(4), 11-35.
- Onwuegbuzie, A. J., & Seaman, M. (1995). The effect of time and anxiety on statistics achievement. *Journal of Experimental Psychology*, 63, 115-124.
- Paterson, C. C. (1996). Self-regulated learning and academic achievement of senior biology students. *Australian Science Teachers Journal*, 42, 48-52.
- Polansky, J., Horan, J. J., & Hanish, C. (1993). Experimental construct validity of the outcomes of study skills training and career counseling as treatment for the retention of at-risk students. *Journal of Counseling and Development*, 71, 488-492.
- Rackliffe, G. (1988, April). *Obstacles to teacher use of the knowledge base for school reform*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Ravid, R., & Leon, M. R. (1995, April). *Students' perceptions of the research component in Master's level teacher education programs*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Sagaria, S. D. (1989, April). Teaching traditional and non-traditional age individuals: How should methods, expectations, and standards differ. Paper presented at the annual meeting of the American Educational Research Association. (ERIC Document Reproduction Service No. ED 308 363)
- Sarason, I. G. (1972). Experimental approaches to test anxiety: Attention and the uses of information. In C. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 2). New York: Academic Press.
- Sen, A. K., & Srivastava, M. (1990). *Regression analysis: Theory, methods, and applications*. New York: Springer-Verlag.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test, for normality and complete samples. *Biometrika*, 52, 592-611.
- Shapiro, S. S., Wilk, M. B., & Chen, H. J. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association*, 63, 1343-1372.
- Slate, J. R., Jones, C. H., & Harlan, E. J. (1998). Study skills of students at a post-secondary vocational technical-institute. *Journal of Industrial Teacher Education*, 35, 57-68.
- Thompson, B. (1994). The pivotal role of replication in psychological research: Empirically, evaluating the replicability of sample results. *Journal of Personality*, 62, 157-176.
- Thompson, B. (1998, April). *Five methodology errors in educational research: The pantheon of statistical significance and other faux pas*. Invited address presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement*, 55, 525-534.
- Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. *Educational and Psychological Measurement*, 45, 203-209.
- Tobias, S. (1977). A model for research on the effect of anxiety on instruction. In Sieber, J., O'Neil, H.F., Jr., & Tobias, S. *Anxiety, learning and instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tobias, S. (1980). Math anxiety: What you can do about it. *Today's Education*, 69, 26-29.
- Tomazic, T. J., & Katz, B. M. (1988, August). *Statistical anxiety in introductory applied statistics*. Paper presented at the annual meeting of the American Statistical Association, New Orleans, LA.
- Ware, M. E., & Chastain, J. D. (1989, April). *Person variables contributing to success in introductory statistics*. Paper presented at the annual meeting of the Southwestern Psychological Association. (ERIC Document Reproduction Service No. ED 309 927)
- Williams, J. E. (1996a, April). *Promoting rural students' academic achievements: An examination of self-regulated learning strategies*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY. (ERIC Document Reproduction Service No. ED 396 890)
- Williams, J. E. (1996b, April). The relation between efficacy for self-regulated learning and domain-specific academic performance, controlling for test anxiety. *Journal of Research and Development in Education*, 29, 77-80.
- Wilson, V. (1997, November). *Factors related to anxiety in the graduate statistics classroom*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis, TN.
- Wine, J. (1980). Cognitive-attentional theory of test anxiety. In I.G. Sarason (Ed.), *Test Anxiety: Theory, research and applications*, pp. 349-385. Hillsdale, NJ: Erlbaum.
- Zeidner, M. (1991). Statistics and mathematics anxiety in social science students--some interesting parallels. *British Journal of Educational Psychology*, 61, 319-328.

JOURNAL SUBSCRIPTION FORM

This form can be used to subscribe to *RESEARCH IN THE SCHOOLS* without becoming a member of the Mid-South Educational Research Association. It can be used by individuals and institutions.

Please enter a subscription to *RESEARCH IN THE SCHOOLS* for:

Name: _____

Institution: _____

Address: _____

COST

Individual Subscription (\$25 per year) Number of years _____

Institutional Subscription (\$30 per year) Number of years _____

Foreign Surcharge (\$25 per year, applies to both individual and institutional subscriptions)	Number of years
---	-----------------

Back issues for Volumes 1 - 6 (\$30 per Volume)	Number of Volumes
---	-------------------

TOTAL COST: _____

MAKE CHECKS PAYABLE TO MSERA

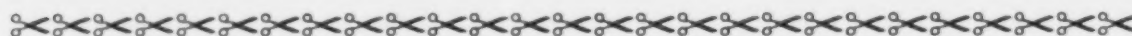
SEND FORM AND CHECK TO:

Dr. James E. McLean, Co-Editor
RESEARCH IN THE SCHOOLS
University of Alabama at Birmingham
School of Education, 233 Educ. Bldg.
901 13th Street, South
Birmingham, AL 35294-1250

MID-SOUTH EDUCATIONAL RESEARCH ASSOCIATION

The Mid-South Educational Research Association (MSERA) was founded in order to encourage quality educational research in the Mid-South and to promote the application of quality educational research in schools. Members of MSERA share interests in educational research, development, and evaluation. While most members are from institutions of higher education, many others represent state departments of education, public and private agencies, and public school systems. Graduate students comprise a significant portion of the membership. A majority of MSERA members are from the six states represented by the organization, but others are from many other states and several foreign countries. The MSERA is the largest regional educational research organization in the country.

The organization provides several services for its members. The annual meeting, held every November, offers many formal and informal opportunities for professional development through special training courses, sharing of research findings, and programmatic interests with colleagues. Members receive a subscription to *RESEARCH IN THE SCHOOLS* and the *Mid-South Educational Researcher*. The MSERA also provides recognition and cash rewards for its outstanding paper, an outstanding dissertation, and professional service.



MSERA Membership/Renewal Form (Please print or type)

Name _____

Organization _____

Address _____

Telephone Work: _____

Home: _____

Fax: _____

e-mail: _____

Amount Enclosed:	MSERA 2000 Membership (\$25 professional, \$15 student)	\$ _____
	MSER Foundation Contribution	\$ _____
	TOTAL	\$ _____

Make check out to MSERA and mail to:

Dr. Robert Calvery
Flippin School District
P. O. Box 239
Flippin, Arkansas 72634

RESEARCH IN THE SCHOOLS
Mid-South Educational Research Association
and the University of Alabama at Birmingham
901 South 13th Street, Room 233
Birmingham, AL 35294-1250

BULK RATE
U.S. POSTAGE
PAID
PERMIT NO. 1256
BIRMINGHAM, AL



RESEARCH IN THE SCHOOLS

A nationally refereed journal sponsored by the
Mid-South Educational Research Association
and the University of Alabama at Birmingham.

Volume 7, Number 2

Fall 2000

Editorial: Statistical Significance Testing and Other Changes to <i>RESEARCH IN THE SCHOOLS</i>	1
<i>James E. McLean and Alan S. Kaufman</i>	
Interpreting Statistical Significance Test Results: A Proposed New "What If" Method	3
<i>Bruce Thompson and Kevin M. Kieffer</i>	
"First We Read a Great Book": Interfacing Literacy Instruction with the Arts in Two Urban Elementary Schools: Preservice Teachers' Case Quandaries and Accomplishments	11
<i>Janet C. Richards and Joan P. Gipe</i>	
A Critique of Methods Used to Answer the Question: Does Money Make a Difference?	21
<i>Jim C. Fortune and Susan Spofford-Richardson</i>	
An Analysis of Factors Related to School-District Size in Arkansas	33
<i>Michael B. Gilbert</i>	
Adolescent Problems and Coping Strategies: Alternative School Versus Non-alternative School Students	41
<i>Deborah D. Vaughn, Ellen K. Slicker, and Judith Van Hein</i>	
High-Stakes Accountability and Contextual Effects: An Empirical Study of the Fairness Issue.	49
<i>Edward B. Reeves</i>	
Developing an Identity of Third-Graders.	59
<i>Sharon L. Newbill and Andrea Clements</i>	
Measures of Work Culture in High and Low Performance Schools	71
<i>Darlene Y. Bruner And Bobbie Jean Greenlee</i>	

RESEARCH IN THE SCHOOLS

Information for Authors

Statement of Purpose

RESEARCH IN THE SCHOOLS (ISSN 1085-5300) publishes original contributions in the following areas: 1) *Research in Practice*--empirical studies focusing on the results of applied educational research including cross-cultural studies, 2) *Topical Articles*--scholarly reviews of research, perspectives on the use of research findings, theoretical articles, and related articles, 3) *Methods and Techniques*--descriptions of technology applications in the classroom, descriptions of innovative teaching strategies in research/measurement/statistics, evaluations of teaching methods, and similar articles of interest to instructors of research-oriented courses, 4) *Assessment*--empirical studies of norm-referenced, criterion-referenced, and informal tests in the areas of cognitive ability, academic achievement, personality, vocational interests, neuropsychological functioning, and the like, and 5) *Other* topics of interest to educational researchers. *RESEARCH IN THE SCHOOLS* is devoted to research conducted in *any* educational setting from a conventional elementary school or high school to a training program conducted within an industry. Likewise, there are no age restrictions on the sample, since the educational settings may include preschools, continuing education classes for adults, or adaptive skills courses in nursing homes. Studies conducted in settings such as clinics, hospitals, or prisons are ordinarily inappropriate for *RESEARCH IN THE SCHOOLS* unless they involve an educational program within such a setting. One goal of *RESEARCH IN THE SCHOOLS* is to provide a training ground for graduate students to learn effective reviewing techniques. Consequently, the journal utilizes a Graduate Student Editorial Board composed mostly of students in educational psychology and educational research. Members of this Editorial Board, each sponsored by a professor, provide supplementary reviews for a selection of submitted articles, and receive both direct and indirect feedback of the quality of these reviews.

Preparing Manuscripts

Authors should prepare manuscripts in accordance with the stylistic rules and guidelines delineated in the *Publications Manual of the American Psychological Association* (4th ed., 1994), which is available from: Order Department, American Psychological Association, PO Box 2710, Hyattsville, MD 20784. Number the pages consecutively. All manuscripts will be subject to editing for sexist language.

Author Identification

Authors should put the complete title of the article on the first text page, but they should exclude their names. Subsequent pages should include only a running head. They should prepare a separate sheet with the complete title of the article and their names and affiliations; this procedure will ensure anonymity in the review process. Authors should supply addresses and phone numbers, and electronic mail addresses and fax numbers (if available), for potential use by the editorial staff and, later, by the production staff. Unless otherwise stated, the first-named author will be sent correspondence, galley proofs, copyright forms, and so forth.

Submission of Manuscripts

Submit manuscripts in triplicate to James E. McLean, Co-Editor, *RESEARCH IN THE SCHOOLS*, Warf-Pickle Hall, Room 418, Eastern Tennessee State University, Box 70685, Johnson City, TN 37614-0685. Please direct questions to jmclean@etsu.edu. All copies should be clear and readable; dot matrix is acceptable only if it meets these qualities of legibility. Length of the manuscripts, including references and tables, should ordinarily range from about 10 to 40 typed, double-spaced, 8-1/2 X 11-inch pages, using 11-12 point type. Abstracts are limited to 125 words. Brief reports of research are not encouraged. Authors are encouraged to keep a hard copy of the manuscript to guard against loss. It is assumed that all manuscripts submitted for publication are original material and have not been simultaneously submitted for publication elsewhere. When manuscripts are accepted for publication, authors are encouraged to submit the final version on a computer disk along with the hard copy.

Copyright and Permissions

Authors are granted permission to reproduce their own articles for personal use. Others must request permission to reproduce tables, figures, or more than 500 words of text from the editors. Copyright © 2000 by the Mid-South Educational Research Association.

EDITORS

James E. McLean, *University of Alabama at Birmingham*
Alan S. Kaufman, *Yale University School of Medicine*

PRODUCTION EDITOR

Margaret L. Rice, *The University of Alabama*

COPY EDITOR

Michele G. Jarrell, *University of Alabama at Birmingham*

EDITORIAL BOARD

Jan E. Alm, *Uppsala University (Sweden)*
Charles M. Achilles, *Eastern Michigan University*
Margery Arnold, *Trinity University*
Mark Baron, *University of South Dakota*
Alan W. Bruc, *Carrollton City (Georgia) Schools*
Robin Cook, *Wichita State University*
Larry G. Daniel, *University of North Texas*
James M. Ernest, *State University of New York at Buffalo*
Daniel Fasko, Jr., *Morehead State University*
Glennelle Halpin, *Auburn University*
Harrison Kane, *University of Nevada, Las Vegas*
JinGyu Kim, *National Institute of Professional Administration (Korea)*
Jwa K. Kim, *Middle Tennessee State University*
Frank R. Lawrence, *University of Alabama at Birmingham*
Robert E. Lockwood, *Alabama State Department of Education*
Robert Marsh, *Chattanooga State Tech. Community College*
Jerry Mathews, *Auburn University*
Charles McLafferty, Jr., *University of Alabama at Birmingham*
Peter C. Melchers, *University of Cologne (Germany)*
Claire Meljac, *Hospital Sainte-Anne (France)*
Soo-Back Moon, *Catholic University of Hyosung (Korea)*
David T. Morse, *Mississippi State University*
Jack A. Naglieri, *Ohio State University*
Anthony J. Onwuegbuzie, *Valdosta State University*
Cecil R. Reynolds, *Texas A & M University*
Janet C. Richards, *The University of Southern Mississippi*
Michael D. Richardson, *Georgia Southern University*
Dan Robinson, *University of Texas at Austin*
John R. Slate, *University of Texas at El Paso*
Paula Stanley, *Radford University*
Bruce Thompson, *Texas A & M University*

GRADUATE STUDENT EDITORIAL BOARD

Keri L. Altig, *University of Nevada, Las Vegas*
Adam S. Bristol, *Yale University*
James M. Graham, *Texas A & M University*
Debbie Hahs, *The University of Alabama*
Keith Hyatt, *University of Nevada, Las Vegas*
James C. Kaufman, *Yale University*
Kevin M. Kieffer, *Texas A & M University*
Ginny Lane, *University of North Texas*
Lynn C. Minor, *Valdosta State University*
William T. Pepper, *Mississippi State University*
Catherine A. Stacy, *University of Texas at Austin*

BEST COPY AVAILABLE

Editorial

Statistical Significance Testing and Other Changes to RESEARCH IN THE SCHOOLS

James E. McLean and Alan S. Kaufman
Editors, *RESEARCH IN THE SCHOOLS*

Since the *RESEARCH IN THE SCHOOLS* Special Issue on Statistical Significance Testing (McLean & Kaufman, 1998), input has been sought by the editors regarding a policy on the reporting of statistical significance testing in the journal. Special sessions have been held at the MSERA Annual Meeting (McLean, November, 1998), requests for input have been included in the journal (Kaufman, 1998), and the editors have sought information from other editors and other sources. The Publication Manual of the American Psychological Association (1994), the style guide required by *RESEARCH IN THE SCHOOLS*, provides little guidance. The Manual discusses both statistical significance (pp. 17-18) and effect size (p. 18), but only "encourages" (p. 18) authors to provide effect size information.

Since the Special Issue, we have also "encouraged" authors to provide effect size information. In fact, we have required authors to provide effect size information to accompany statistical significance tests unless they could provide a compelling reason not to. Since that time, we have had no author make a compelling case to omit effect size information. Thus, we have decided to make this policy explicit and require that authors accompany the reporting of statistical significance tests with effect size information. Specifically, the following line has been added to the *RESEARCH IN THE SCHOOLS* "Information for Authors" section: "All reporting of statistical significance must include an estimate of effect size."

We are hopeful that this change will encourage educational researchers to consider effect size and practical significance when evaluating the results of a study. We are also encouraging our Editorial Board members to consider effect size and the practical significance of a study in their recommendations. In the end, we hope this change supports the movement towards the reporting of more complete and accurate results of research studies. Since this change in policy merely formalizes what we have been practicing for at least two years, we do not expect that it will have an impact on the number of manuscripts we receive, but we do hope it will have an impact on the quality of the manuscripts. We continue to encourage readers' responses on this and any other policy or aspect of *RESEARCH IN THE SCHOOLS*.

We would also like to share some other exciting news about the journal. The manuscript load has increased to such an extent that we have decided to add two associate editors. In keeping with our practice of combining seasoned veterans with up-and-coming newcomers, we have asked Dr. Charles M. Achilles of Eastern Michigan University and Dr. James C. Kaufman of the Educational Testing Service to serve as associate editors and they both have accepted.

As a consequence of the success of our first special issue on statistical significance testing, we have decided to inaugurate our new associate editors by assigning them the primary responsibility for two new special issues. Dr. Achilles will provide the leadership for a special issue to focus upon how schools and classrooms are organized for providing education services to students. Topics might include, but not be limited to, class size, use of para-professionals, non-graded or multi-grades schools, privatization (Charters, vouchers, etc.), school size. Dr. James Kaufman will be the lead editor for a special issue on "Creativity in the Schools." In this issue, we will be asking several experts in creativity and education to contribute a short piece on their views of the following questions. What is the current state of creativity in education? Should there be more of an attempt made to teach and nurture student creativity? If so, then how can this goal be balanced with other educational goals? What new theories and research on creativity should be incorporated into the classroom? Is creativity a "luxury" for the gifted classes, or should teachers try to share it with mainstream classes as well? How about those with learning disabilities or mental retardation? These special issues will appear in 2002 and 2003.

References

- American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
- Kaufman, A. S. (1998). Introduction to the special issue on statistical significance testing. *RESEARCH IN THE SCHOOLS*, 5(2), 1.

McLean, J. E. (1998, November). *Statistical significance testing—What should Research in the Schools' position be?* Session presented at the annual meeting of the Mid-South Educational Research Association in New Orleans, LA.

McLean, J. E., & Kaufman, A. S. (Editors). Statistical significance testing [Special Issue]. *RESEARCH IN THE SCHOOLS*, 5(2).

Interpreting Statistical Significance Test Results: A Proposed New "What If" Method

Bruce Thompson
Texas A&M University
and
Baylor College of Medicine

Kevin M. Kieffer
Tampa VA Medical Center

As the 1994 APA publication manual emphasized, p values are affected by sample size. Thus, it can be helpful to interpret the results of statistical significance tests in a sample size context by conducting so-called "what if" analyses. However, these methods can be inaccurate unless "corrected" effect sizes are employed. The present paper proposes a new method by which "what if" analyses can be conducted using estimated true population effects.

A recent empirical study of journals published since 1950 in education, psychology, medicine, and ecology shows a geometric decade-by-decade growth in the number of articles criticizing statistical significance testing (Anderson, Burnham, & Thompson, 1999). The related controversy has even led to a special theme issue of the MSERA journal, *RESEARCH IN THE SCHOOLS* (cf. Daniel, 1998; McLean & Ernest, 1998; Nix & Barnette, 1998). Following scheduled debates at the annual meetings of both the American Psychological Association (APA) and the American Psychological Society, in 1996 the APA Board of Scientific Affairs appointed its Task Force on Statistical Inference (Azar, 1997, 1999; Shea, 1996). The wide-ranging, comprehensive and thoughtful report of the Task Force was published in the August 1999 issue of the *American Psychologist* (Wilkinson & The APA Task Force on Statistical Inference, 1999).

In their historical summary dating back to the origins of these tests, Huberty and Pike (1999) provide a thoughtful review of how we got to where we are as regards statistical tests. Among the recent articles criticizing statistical testing practices, Cohen (1994), Kirk (1996), Rosnow and Rosenthal (1989), Schmidt (1996), and

Thompson (1996) have been especially thoughtful. However, these criticisms are certainly not new (see Boring, 1919). Among the classical criticisms, Carver (1978), Meehl (1978), and Rozeboom (1960) in particular have been widely cited.

Among the more thoughtful works advocating conventional statistical testing, Cortina and Dunlap (1997), Frick (1996), and especially Abelson (1997), have been most influential. A balanced and comprehensive treatment of both perspectives is provided by Harlow, Mulaik and Steiger (1997) (for reviews of this book, see Levin, 1998 and Thompson, 1998).

Tenor of More Extreme Conclusions

Two quotations may convey the tenor of some of the conclusions of some more extreme critics of statistical tests. Rozeboom (1997) recently argued that

Null-hypothesis significance testing is surely the most bone-headedly misguided procedure ever institutionalized in the rote training of science students . . . [I]t is a sociology-of-science wonderment that this statistical practice has remained so unresponsive to criticism . . . (p. 335)

And Tryon (1998) recently lamented,

[T]he fact that statistical experts and investigators publishing in the best journals cannot consistently interpret the results of these analyses is extremely disturbing. Seventy-two years of education have resulted in minuscule, if any, progress toward correcting this situation. It

Bruce Thompson is Professor and Distinguished Research Scholar, Texas A&M University, and Adjunct Professor of Community Medicine, Baylor College of Medicine. Kevin M. Kieffer is an intern at the James Haley VA Medical Center in Tampa, and a doctoral candidate at Texas A&M University. The senior author may be e-mailed (and related reprints accessed) through Web URL: "<http://acs.tamu.edu/~bbt6147/>". The junior author may be contacted at: "kmkieffer@earthlink.net".

is difficult to estimate the handicap that widespread, incorrect, and intractable use of a primary data analytic method has on a scientific discipline, but the deleterious effects are doubtless substantial . . . (p. 796)

Indeed, *empirical* studies confirm that many researchers do not fully understand the logic of their statistical tests (cf. Mittag, 1999; Nelson, Rosenthal & Rosnow, 1986; Oakes, 1986; Rosenthal & Gaito, 1963; Zuckerman, Hodgins, Zuckerman & Rosenthal, 1993). Misconceptions are taught even in widely-used statistics textbooks (Carver, 1978).

Purpose of the Present Paper

The recent fourth edition of the American Psychological Association style manual (APA, 1994) explicitly acknowledged that *p* values are not acceptable indices of effect:

Neither of the two types of [statistical significance] probability values reflects the importance or magnitude of an effect *because both depend on sample size* . . . You are [therefore] encouraged to provide effect-size information. (APA, 1994, p. 18, emphasis added)

Indeed, the author guidelines for *Measurement and Evaluation in Counseling and Development* encourage authors

. . . to assist readers in interpreting statistical significance of their results. For example, results may be indexed to sample size. An author may wish to say, "this correlation coefficient would have still been statistically significant even if sample size had been as small as $n = 33$," or "this correlation coefficient would have been statistically significant if sample size had been increased to $n = 138$." (Association for Assessment in Counseling, 1994, p. 143)

Thompson (1989a, 1989b) proposed methods for conducting such "what if" analyses, which help researchers interpret their results by considering the extent to which sample size (as against effect size) yielded statistical significance.

The purpose of the present paper is to propose a new method for conducting such "what if" analyses to augment the conventional use of statistical significance tests. It will

momentarily be noted that the methods previously proposed (Thompson, 1989a, 1989b) have some serious weaknesses, which the new proposed methods address and overcome.

Sample Size Influences

When "nil" null hypotheses specifying zero relationship or zero group differences (Cohen, 1994) are used, the null will always be rejected at some sample size. There are infinitely many possible sample effects (Kirk, 1996), and therefore the probability of obtaining an exactly zero sample effect, as specified by a "nil" null, is infinitely small. That is, because probability equals the given occurrence divided by the total number of occurrences, given that the denominator is infinite, the probability is infinitely small. Therefore, given a "nil" null, and a non-zero sample effect, the null hypothesis will *always* be rejected at some sample size!

As Hays (1981) emphasized, "virtually any study can be made to show significant results if one uses enough subjects" (p. 293). This means that

Statistical significance testing can involve a tautological logic in which tired researchers, having collected data from hundreds of subjects, then conduct a statistical test to evaluate whether there were a lot of subjects, which the researchers already know, because they collected the data and know they're tired. (Thompson, 1992, p. 436)

Certainly this dynamic is well known, even though its influence is just as widely underestimated. Table 1 illustrates how sample size impacts $p_{\text{CALCULATED}}$ values for a hypothetical one-way four-level ANOVA study (or, alternatively, a regression study involving three predictor variables). These results involve an η^2 value, 13.8%, that Cohen (1988, pp. 26-27) characterized as "large" as regards result typicality. [In an ANOVA, η^2 is the percentage of variance in the dependent variable that can be predicted with knowledge of the participants' memberships in the study's groups or design cells.]

More to the point, statistical significance testing with "nil" null hypotheses is arguably *irrelevant either when (a) sample size is very large or (b) effect size is very large*. Table 2 presents illustrative results for the Pearson product-moment correlation coefficient (including bivariate reliability or validity coefficients). The first three effect sizes were those characterized by Cohen (1988, pp. 24-27) as "low," "medium," and "large," as regards result typicality.

A NEW "WHAT IF" METHOD

Table 1
Illustrative Variations in p for a Fixed Effect Size
Across Various Sample Sizes

n	Statistic			
	η^2	ω^2	F_{calc}	p_{calc}
6	13.8%	-80.7%	0.11	.9487
12	13.8%	-13.6%	0.48	.7040
18	13.8%	-3.3%	0.80	.5127
24	13.8%	1.4%	1.12	.3632
30	13.8%	4.1%	1.44	.2528
36	13.8%	5.8%	1.76	.1738
42	13.8%	7.0%	2.08	.1184
48	13.8%	7.9%	2.40	.0801
54	13.8%	8.6%	2.72	.0539
60	13.8%	9.1%	3.04	.0361
66	13.8%	9.6%	3.36	.0241
72	13.8%	9.9%	3.68	.0160
78	13.8%	10.2%	4.00	.0106
84	13.8%	10.5%	4.32	.0070
90	13.8%	10.7%	4.64	.0047
96	13.8%	10.9%	4.96	.0031
102	13.8%	11.1%	5.28	.0020

Note. The ANOVA effect size η^2 (η^2) is an "uncorrected" variance-accounted-for statistic that characterizes sample results in a manner analogous to r^2 ; this effect does not consider (and therefore does not vary) over changes in sample size. However, Hays' ω^2 (ω^2) is a "corrected" variance-accounted-for estimate of population effects that does take into consideration those study features (e.g., sample size) that create more or less positive bias in the sample effect size (see Snyder & Lawson, 1993). Of course, both F and p calculations also take into account sample size, and thus vary even for a fixed sample effect over changes in sample size.

Table 2
Illustrations of How Statistically Significant Results
Occur Either with Large n 's and Small Effects or
Large Effects and Small n 's

n	Statistic			
	η^2	ω^2	F_{calc}	p_{calc}
385	1.0%	0.7%	3.87	.0499
65	5.9%	4.4%	4.01	.0494
28	13.8%	10.3%	4.32	.0472
4	70.0%	54.5%	7.00	.0773
5	70.0%	58.1%	9.33	.0378
3	80.0%	63.6%	8.00	.1056
4	80.0%	68.8%	12.00	.0405
3	90.0%	81.0%	18.00	.0513
4	90.0%	83.9%	27.00	.0138

Note. Sample sizes (n) and p values less than .05 are presented in bold. The first three effect sizes were those characterized by Cohen (1988, pp. 24-27) as "low," "medium," and "large," as regards result typicality.

More than 60 years ago, Berkson (1938) wrote an article titled, "Some difficulties of interpretation encountered in the application of the chi-square test." He noted that when working with data from roughly 200,000 people,

an observant statistician who has had any considerable experience with applying the chi-square test repeatedly will agree with my statement that, as a matter of observation, when the numbers in the data are quite large, the P 's tend to come out small . . . [W]e know in advance the P that will result from an application of a chi-square test to a large sample . . . But since the result of the former is known, it is no test at all! (pp. 526-527)

Some 30 years ago, Bakan (1966) reported that, "The author had occasion to run a number of tests of significance on a battery of tests collected on about 60,000 subjects from all over the United States. Every test came out significant" (p. 425). Shortly thereafter, Kaiser (1976) reported not being surprised when many substantively trivial factors were found to be statistically significant when data were available from 40,000 participants.

Particularly egregious is the use of "nil" nulls to test measurement hypotheses, where wildly non-nil results are both anticipated and demanded. As Abelson (1997) explained,

And when a reliability coefficient is declared to be nonzero, that is the ultimate in stupefyingly vacuous information. What we really want to know is whether an estimated reliability is .50'ish or .80'ish. (p. 121)

Thus, Table 2 also illustrates that reliability or validity coefficients of this magnitude will be statistically significant even with n 's as small as four or five people.

The Old "What If" Method

Table 3 illustrates the use of the "what if" method described by Thompson (1989a, 1989b). In this method certain values are taken as fixed (presented in bold in the table). The table presumes a regression study involving two predictor variables ($df_{EXPLAINED} = 2$) and $n = 40$. The first three rows in the table present hypothetical results from this study.

Then the sample size is varied, assuming a fixed effect size (e.g., R^2 , η^2). [The tabled results here could equally well be viewed as a three-level one-way ANOVA problem ($df_{EXPLAINED} = k-1 = 2$), because $R^2 =$

$SOS_{EXPLAINED} / SOS_{TOTAL}$, but also $\eta^2 = SOS_{EXPLAINED} / SOS_{TOTAL}$.] In the Table 3 example, given the design and a fixed variance-accounted-for effect size of 10.0%, the effect size becomes statistically significant ($\alpha=.05$) when n goes from 59 to 60. These analyses can be easily conducted using a microcomputer spreadsheet such as Excel; Appendix A presents the commands.

Table 3
Illustration of Previously Proposed (Thompson, 1989a, 1989b)
"What If" Analyses

Source	SOS	df	MS	F_{calc}	p_{calc}	R^2	Adj R^2
$n = 40$							
Exp	10.000	2	5.000	2.056	.142392	10.00%	2.50%
Unexp	90.000	37	2.432				
Total	100.000	39	2.564				
$n = 59$							
Exp	10.000	2	5.000	3.111	.052335	10.00%	5.09%
Unexp	90.000	56	1.607				
Total	100.000	58	1.724				
$n = 60$							
Exp	10.000	2	5.000	3.167	<u>.049649</u>	10.00%	5.18%
Unexp	90.000	57	1.579				
Total	100.000	59	1.695				

Note. Fixed values in the "what if" analysis are presented in bold.

A New Proposed Method

"Corrected" vs "Uncorrected" Effect Sizes

"Classical" statistical methods (e.g., ANOVA, regression) use the statistical theory called "ordinary least squares." This theory optimizes the fit of the synthetic/latent variables (e.g., $\hat{Y}$) to the observed/measured outcome/response variables (e.g., Y) in the sample data, and capitalizes on *all* the variance present in the observed sample scores, including the "sampling error variance" that is idiosyncratic to the particular sample. Because sampling error variance is unique to a given sample (i.e., each sample has its own sampling error variance), "uncorrected" variance-accounted-for effect sizes (e.g., R^2 , η^2) somewhat overestimate the effects that would be replicated in either (a) the population or (b) a future sample.

However, statistical theory can be invoked to estimate the extent of overestimation (i.e., positive bias) in the sample variance-accounted-for effect size estimate. [Note that "corrected" estimates are always less than or equal to "uncorrected" values.] The difference between the "uncorrected" sample (e.g., R^2 , η^2) and "corrected" population (e.g., "adjusted R^2 ," Hays' ω^2) variance-accounted-for effect sizes is called "shrinkage." That is, these "corrected" effect size estimates are estimates of the effect size in the population.

For example, for regression the "corrected" effect size "adjusted R^2 " is automatically provided by most statistical packages, even without being requested. This correction is due to Ezekiel (1930), although the formula is often incorrectly attributed to Wherry (Kromrey & Hines, 1996):

$$1 - ((n - 1) / (n - v - 1)) * (1 - R^2),$$

where n is the sample size and v is the number of predictor variables. For example, if $n = 60$, and there are six predictor variables in a regression analysis for which $R^2 = 50.0\%$, "adjusted R^2 "

$$\begin{aligned} &= 1 - [(n - 1) / (n - v - 1)] * (1 - R^2) \\ &= 1 - [(60 - 1) / (60 - 6 - 1)] * (1 - .5) \\ &= 1 - [(59) / (60 - 6 - 1)] * (1 - .5) \\ &= 1 - [(59) / (53)] * (1 - .5) \\ &= 1 - [(59) / (53)] * (.5) \\ &= 1 - [1.113] * (.5) \\ &= 1 - 0.556 \\ &= .44339 = 44.3\%. \end{aligned}$$

The "adjusted R^2 " formula can also be equivalently expressed as:

$$R^2 - ((1 - R^2) * (v / (n - v - 1))).$$

In the ANOVA case, the analogous ω^2 can be computed using the formula due to Hays (1981, p. 349):

$$(SS_{BETWEEN} - (k - 1) * MS_{WITHIN}) / (SS_{TOTAL} + MS_{WITHIN}),$$

where k is the number of groups.

Problem with the Previous (Thompson, 1989a, 1989b) Method

The previously proposed "what if" methods (Thompson, 1989a, 1989b), as illustrated in Table 3, are problematic in that they are conducted in the metric of the sample (i.e., an "uncorrected" effect size) rather than in the metric of the population from which all samples are ostensibly drawn (i.e., a "corrected" effect size). Put differently, *the previous methods do not take into account that the amount of sampling error (and therefore the positive bias in the "uncorrected" effect size) will change as sample size itself changes.*

Proposed Alternative "What If" Method

The alternative "what if" analytic method proposed here invokes the "corrected" estimate of the *population* effect size as the metric for exploring sample size influences. This analysis is illustrated in Table 4 using the same research design and results presumed in the first three rows of Table 3. Appendix B presents the spreadsheet commands that readily implement the proposed analyses.

A NEW "WHAT IF" METHOD

Table 4
Illustration of New Proposed "What If" Methods Using
Fixed "Corrected" Effect Size

Source	SOS	df	MS	F_{calc}	p_{calc}	R^2	Adj R^2
$n = 40$							
Exp	10.000	2	5.000	2.056	.142392	10.00%	2.50%
Unexp	90.000	37	2.432				
Total	100.000	39	2.564				
$n = 60$							
Exp	7.458	2	3.729	2.297	.109827	7.46%	2.50%
Unexp	92.542	57	1.624				
Total	100.000	59	1.695				
$n = 80$							
Exp	6.203	2	3.101	2.546	.084988	6.20%	2.50%
Unexp	93.797	77	1.218				
Total	100.000	79	1.266				
$n = 121$							
Exp	4.937	2	2.469	3.064	.050413	4.94%	2.50%
Unexp	95.063	118	.806				
Total	100.000	120	.833				
$n = 122$							
Exp	4.917	2	2.459	3.077	.049777	4.92%	2.50%
Unexp	95.083	119	.799				
Total	100.000	121	.826				

Note. Fixed values in the "what if" analysis are presented in bold.

In these analyses "adjusted R^2 " (or ω^2) is taken as fixed, while R^2 (or η^2) changes along with sample size variations. Given "adjusted R^2 ," R^2 can be solved for by algebraically rearranging the Ezekiel (1930) correction formula. If, for example, "adjusted R^2 " = 44.3%, $n = 60$, and there are six regression predictor variables, ("uncorrected") R^2

$$\begin{aligned}
 &= 1 - [1 * ((\text{Adj } R^2 - 1) / ((n - 1) / (n - v - 1)))] \\
 &= 1 - [1 * ((.4433 - 1) / ((60 - 1) / (60 - 6 - 1)))] \\
 &= 1 - [1 * ((.4433 - 1) / ((60 - 1) / (53)))] \\
 &= 1 - [1 * ((.4433 - 1) / ((59) / (53)))] \\
 &= 1 - [1 * ((-.556) / ((59) / (53)))] \\
 &= 1 - [1 * ((-.556) / (1.113))] \\
 &= 1 - [1 * (-.5)] \\
 &= 1 - [-.5] \\
 &= .5 = 50.0\%.
 \end{aligned}$$

Discussion

The illustrative results presented in Tables 3 and 4 make clear how different the results can be from the two "what if" analytic strategies. In the illustrations, a researcher obtains a squared multiple correlation coefficient (R^2) of 10.0% in a hypothetical study involving two predictor variables and 40 participants.

Using the classical "what if" analyses proposed by Thompson (1989a, 1989b), the unadjusted effect size for

this design becomes statistically significant ($\alpha = .05$) when n goes from 59 to 60, as reported in Table 3. However, for the same design, using what is suggested here is the more accurate new proposed "what if" analytic strategy, the results become statistically significant ($\alpha = .05$) when n goes from 121 to 122, as reported in Table 4.

Applicability with Results that Were Originally Statistically Significant

The previous discussion has involved a hypothetical research scenario in which the original results were not statistically significant. However, it is emphasized that the proposed strategy is not limited to use with results that were originally statistically non-significant.

The proposed strategy can be just as useful with results that were originally statistically significant, just as the previous "what if" methods could also be employed with just such results (Thompson, 1989a, 1989b). As the author guidelines for *Measurement and Evaluation in Counseling and Development* note, for example,

An author may wish to say, "this [statistically significant] correlation coefficient would have still been statistically significant even if sample size had been as small as $n = 33$ "... (Association for Assessment in Counseling, 1994, p. 143)

Benefits of "What If" Analyses

Use of these "what if" methods may prevent authors with large sample sizes from overinterpreting their small effects, once they see that the small effects would no longer have been statistically significant with only a slightly smaller sample size. Conversely, researchers with large effects will be even more confident in interpreting their results if they note that their observed effects would still have been statistically significant even if they had had an appreciably smaller sample size.

Example Applications for Recently Published Articles

Small sample size, small effect size. Lightsey and Christopher (1997) conducted a study of what variables predicted variability in the depression of 60 participants. For example, they reported that thinking positive thoughts had an r^2 effect size of 5%. The result was not statistically significant ($\alpha = .05$).

However, a "what if" analysis of the results, given the design, and using the proposed methods and the appended spreadsheet commands, indicated that the result would have been statistically significant if only 2 more people had participated in the study, assuming that the effect size would then be roughly the same. The proposed "what if" analysis suggests that the result may still be noteworthy and reminds us that "surely, God loves the .06 nearly as

much as the .05" level of statistical significance (Rosnow & Rosenthal, 1989, p. 1277).

Large Effect Size. In a study of 195 women's perceptions of counselors with one of three different orientations as regards feminism, Hackett, Enns, and Zetzer (1992) reported a three-level one-way ANOVA result for which the $F_{\text{CALCULATED}}$ value was 121.74 ($df=2$, 192). By employing the proposed new methods and appended spreadsheet commands, it was determined that the η^2 value for this result was 55.9%, while the adjusted effect size was 55.2%.

This effect size is several times larger than the effect size that Cohen (1988) characterized as "large." The "what if" analysis of the result makes clear just how large the effect was. The results indicate that this effect size would have remained statistically significant with an n as small as 7 people!

Large Sample Size. For the first major hypothesis that she tested, Voelkl (1995) had a sample size of 13,098 participants for a multiple regression problem involving four predictor variables. She reported that, "The overall test of association between school warmth and four student achievement measures was statistically significant, [sic] $F(4, 13093) = 23.83, p < .0001$ " (p. 133). For these results we can use a computer spreadsheet to determine that R^2 was 0.72% and the actual $p_{\text{CALCULATED}}$ value was $1.2E-19$ (i.e., 19 zeroes to the right of the decimal, followed by "12"). The author's discussion of this result was quite succinct and was offered without any reference to the influence of sample size: "Student perceptions of school warmth were significantly related to academic achievement" (p. 136).

By employing the proposed new methods using the Appendix B spreadsheet commands, we find that here the sample results for a population "adjusted R^2 " (i.e., 0.68%) remains statistically significant until the n drops from 657 to 656. The result clearly indicates that, although the original sample size was huge, a very large sample remains necessary for this very small effect size to remain statistically significant ($\alpha=.05$). This result from the proposed "what if" analysis reinforces the notion that p values *cannot* be used as reasonable indices of effect, and that the results for this hypothesis were not noteworthy even though they were statistically significant.

Caveat

It should be noted that the logic proposed here could be employed with other of the many "corrected" effect sizes presented in the literature (cf. Kirk, 1996; Snyder & Lawson, 1993). Some of the various alternatives are more liberal or more conservative, depending upon whether and the manner in which they estimate population or future sample effects.

Of course, the proposed analyses and related methods (see Morse, 1999) do not constitute either magic or a panacea. But the proposed methods may help researchers to see how their sample size may have impacted their calculated p values.

It is also very important to emphasize that the analyses do not have to presume that an exact effect size would be replicated in a future study with more (or fewer) participants. Indeed, in addition to conducting the analyses illustrated here, it may be very useful to conduct the same analyses with **both** somewhat larger and somewhat smaller effect sizes, so as to model sample size impacts for a given design across a reasonable range of effect size outcomes!

As suggested elsewhere,

In any case, the purpose of this approach is not to identify the exact results that would occur with a different sample size, assuming exactly the same effect size. Rather, the approach focuses on establishing a general ballpark for interpreting statistical significance tests in a sample size context. (Thompson, 1993, p. 368)

Thus, "what if" analysis, just like any other analyses, should not be over-interpreted and must be employed reasonably and reflectively.

References

- Abelson, R.P. (1997). A retrospective on the significance test ban of 1999 (If there were no significance tests, they would be invented). In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), *What if there were no significance tests?* (pp. 117-141). Mahwah, NJ: Erlbaum.
- American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
- Anderson, D.R., Burnham, K.P., & Thompson, W.L. (1999). *Null hypothesis testing in ecological studies: Problems, prevalence, and an alternative*. Manuscript submitted for publication.
- Association for Assessment in Counseling. (1994). Guidelines for authors. *Measurement and Evaluation in Counseling and Development*, 27, 341.
- Azar, B. (1997). APA task force urges a harder look at data. *The APA Monitor*, 28(3), 26.
- Azar, B. (1999). APA statistics task force prepares to release recommendations for public comment. *The APA Monitor*, 30(5), 9.
- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, 66, 423-437.

- Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association*, 33, 526-536.
- Boring, E.G. (1919). Mathematical vs. scientific importance. *Psychological Bulletin*, 16, 335-338.
- Carver, R. (1978). The case against statistical significance testing. *Harvard Educational Review*, 48, 378-399.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.
- Cortina, J.M., & Dunlap, W.P. (1997). Logic and purpose of significance testing. *Psychological Methods*, 2, 161-172.
- Daniel, L.G. (1998). Statistical significance testing: A historical overview of misuse and misinterpretation with implications for the editorial policies of educational journals. *RESEARCH IN THE SCHOOLS*, 5(2), 23-32.
- Ezekiel, M. (1930). *Methods of correlational analysis*. New York: Wiley.
- Frick, R.W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods*, 1, 379-390.
- Hackett, G., Enns, C.Z., & Zetzer, H.A. (1992). Reactions of women to nonsexist and feminist counseling: Effects of counselor orientation and mode of information delivery. *Journal of Counseling and Development*, 39, 321-330.
- Harlow, L.L., Mulaik, S.A., & Steiger, J.H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.
- Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart and Winston.
- Huberty, C.J., & Pike, C.J. (1999). On some history regarding statistical testing. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 1-22). Stamford, CT: JAI Press.
- Kaiser, H.F. (1976). Review of *Factor analysis as a statistical method*. *Educational and Psychological Measurement*, 36, 586-589.
- Kirk, R. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56, 746-759.
- Kromrey, J.D., & Hines, C.V. (1996). Estimating the coefficient of cross-validity in multiple regression: A comparison of analytical and empirical methods. *Journal of Experimental Education*, 64, 240-266.
- Levin, J.R. (1998). To test or not to test H_0 ? *Educational and Psychological Measurement*, 58, 311-331.
- Lightsey, O.W., Jr., & Christopher, J.C. (1997). Stress buffers and dysphoria in a non-western population. *Journal of Counseling and Development*, 75, 451-459.
- McLean, J.E., & Ernest, J.M. (1998). The role of statistical significance testing in educational research. *RESEARCH IN THE SCHOOLS*, 5(2), 15-22.
- Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.
- Mittag, K.G. (1999, April). *A national survey of AERA members' perceptions of the nature and meaning of statistical significance tests*. Paper presented at the annual meeting of the American Educational Research Association, Montreal. (ERIC Document Reproduction Service No. ED 429 086)
- Morse, D.T. (1999). MINSIZE2: A computer program for determining effect size and minimum sample size for statistical significance for univariate, multivariate, and nonparametric tests. *Educational and Psychological Measurement*, 59, 518-531.
- Nelson, N., Rosenthal, R., & Rosnow, R.L. (1986). Interpretation of significance levels and effect sizes by psychological researchers. *American Psychologist*, 41, 1299-1301.
- Nix, T.W., & Barnette, J.J. (1998). The data analysis dilemma: Ban or abandon. A review of null hypothesis significance testing. *RESEARCH IN THE SCHOOLS*, 5(2), 3-14.
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. New York: Wiley.
- Rosenthal, R., & Gaito, J. (1963). The interpretation of level of significance by psychological researchers. *Journal of Psychology*, 55, 33-38.
- Rosnow, R.L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, 44, 1276-1284.
- Rozeboom, W.W. (1960). The fallacy of the null hypothesis significance test. *Psychological Bulletin*, 57, 416-428.
- Rozeboom, W.W. (1997). Good science is abductive, not hypothetico-deductive. In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), *What if there were no significance tests?* (pp. 335-392). Mahwah, NJ: Erlbaum.
- Schmidt, F.L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods*, 1, 115-129.

- Shea, C. (1996). Psychologists debate accuracy of "significance test." *Chronicle of Higher Education*, 42(49), A12, A16.
- Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education*, 61, 334-349.
- Thompson, B. (1989a). Asking "what if" questions about significance tests. *Measurement and Evaluation in Counseling and Development*, 22(2), 66-68.
- Thompson, B. (1989b). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues. *Measurement and Evaluation in Counseling and Development*, 22(1), 2-5.
- Thompson, B. (1992). Two and one-half decades of leadership in measurement and evaluation. *Journal of Counseling and Development*, 70, 434-438.
- Thompson, B. (1993). The use of statistical significance tests in research: Bootstrap and other alternatives. *Journal of Experimental Education*, 61, 361-377.
- Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher*, 25(2), 26-30.
- Thompson, B. (1998). Review of *What if there were no significance tests?* by L. Harlow, S. Mulaik & J. Steiger (Eds.). *Educational and Psychological Measurement*, 58, 332-344.
- Tryon, W.W. (1998). The inscrutable null hypothesis. *American Psychologist*, 53, 796.
- Voelkl, K.E. (1995). School warmth, student participation, and achievement. *Journal of Experimental Education*, 63, 127-138.
- Wilkinson, L., & The APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604. [reprint available through the APA Home Page: <http://www.apa.org/journals/amp/amp548594.html>]
- Zuckerman, M., Hodgins, H.S., Zuckerman, A., & Rosenthal, R. (1993). Contemporary issues in the analysis of data: A survey of 551 psychologists. *Psychological Science*, 4, 49-53.

Appendix A
Excel Spreadsheet Commands
for Previous "What If" Method (Table 3)

Input							
A	B	C	D	E	F	G	H
1 Source	SOS	df	MS	Fcalc	pcalc	R ²	Adj R ²
2 Exp	+b4*g2	2	+b2/c2	+d2/d3	Note F1	10.00%	Note H1
3 Unexp	+b4-b2	+c4-c2	+b3/c3				
4 Total	100.000	+c5-1	+b4/c4				
5	n = 40						

Note. Cell F1 = =fdist(c2,c2,c3)
Cell H1 = =1-(((+c4)/(+c4-(c2+1))))*(1-g2))

Output:							
A	B	C	D	E	F	G	H
1 Source	SOS	df	MS	Fcalc	pcalc	R ²	Adj R ²
2 Exp	10.000	2	5.000	2.056	.142392	10.00%	2.50%
3 Unexp	90.000	37	2.432				
4 Total	100.000	39	2.564				
5	n = 40						

Appendix B
Excel Spreadsheet Commands
for Proposed "What If" Method (Table 4)

Input:							
A	B	C	D	E	F	G	H
1 Source	SOS	df	MS	Fcalc	pcalc	R ²	Adj R ²
2 Exp	+g2*b4	2	+b2/c2	+d2/d3	Note F1	Note G1	2.50%
3 Unexp	+b4-b2	+c4-c2	+b3/c3				
4 Total	100.000	+c5-1	+b4/c4				
5	n = 122						

Note. Cell F1 = =fdist(c2,c2,c3)
Cell G1 = =1-(-1*((h2-1)/((c4)/(c4-(c2+1))))))

Output:							
A	B	C	D	E	F	G	H
1 Source	SOS	df	MS	Fcalc	pcalc	R ²	Adj R ²
2 Exp	4.917	2	2.459	3.077	.049777	4.92%	2.50%
3 Unexp	95.083	119	.799				
4 Total	100.000	121	.826				
5	n = 122						

**"First We Read a Great Book":
Interfacing Literacy Instruction with the Arts in Two Urban Elementary Schools:
Preservice Teachers' Case Quandaries and Accomplishments**

Janet C. Richards

University of Southern Mississippi

Joan P. Gipe

University of New Orleans and California State University, Sacramento

This qualitative inquiry examines preservice teachers' cases describing their reflective thinking, hopes, concerns, problems, and achievements as they address the complexities of devising and offering literacy-based arts lessons to students in two urban, elementary schools. The researchers examined and categorized 203 teaching cases written by eight groups of preservice teachers. Analysis revealed 11 major themes. Ten themes related to problematic issues, such as classroom teachers' apathy and disinterest in the arts and preservice teachers' lack of modeling and scaffolding during arts lessons. A single theme highlighted the achievements and successes of the preservice teachers as they recognized the increased motivation and enthusiasm of disruptive or academically at-risk students when students became fully engaged in arts activities. Implications of the inquiry underscores the importance of providing preservice teachers with hands-on encounters with the arts and the potential dangers of reducing or eliminating the arts in elementary education.

"The trouble began when Richard refused to practice his part. He wanted to make up his own lines instead of following the script I wrote" (excerpt from a preservice teacher's case narrative)

"Well, she sang like a bird! Her creative book was perfect ... and she remembered the entire story of *The Little Red Hen*. Who would have thought that her behavior would improve so dramatically when I introduced the arts?" (excerpt from a preservice teacher's case narrative)

"These third graders never have been allowed to manipulate the medium of paint so they mixed all the paint together until they formed a greyish-green color instead of the red and yellow we needed for our pyramid" (excerpt from a preservice teacher's case narrative)

"I ain't gonna be no little pig in a stupid play," Alan shouted. "That's fine," I said. "You can just stay up here in the classroom with your teacher" (excerpt from a preservice teacher's case narrative)

While there have been nominal attempts to acknowledge the arts in official statements of United States educational goals, the arts are neglected in *Goals 2000: The Educate America Act* (1994) (Greene, 1995). One possible explanation for this oversight is the reality that artistic pursuits remain peripheral in preservice teacher education because "they exist outside the framework of what the elementary teacher is required to teach and is truly held accountable for" (Fowler, 1988, p. 55). Not surprisingly, a survey reveals that United States' teachers overwhelmingly believe that arts activities are time-consuming 'frills,' with little practical use (Bolton, 1985). A few scholars even suggest that teachers who wish to offer literacy-based arts activities might first consider participating "in these exercises under the direction of a trained professional" (Popp, 1996, p. 235). Unfortunately, "only one in four students [in United States schools] gets the chance to sing, play an instrument or perform plays in class each week" (Mississippi *Sun Herald*, November 11, 1998, Section B, p. 1), and Richard Riley, United States Secretary of Education, notes a continuing retreat of the arts in schools (Mississippi *Sun Herald*, November 12, 1998, Section B, p. 4).

An earlier version of this article was presented at the 11th European Conference on Reading, Stavanger, Norway, August, 1999. Janet C. Richards is Professor in the College of Education at the University of Southern Mississippi, Long Beach. Joan P. Gipe is Professor Emeritus of Education at the University of New Orleans and Lecturer at the California State University, Sacramento. Correspondence regarding this paper should be addressed to Janet C. Richards, University of Southern Mississippi, 730 East Beach Blvd., Long Beach, MS 39560 or via e-mail at janetum@aol.com.

How We View the Arts

As longtime supervisors of two university/urban elementary school literacy collaborations, we view the arts "not as extras to be addressed if there is time left over in the day" (Blecher & Jaffee, 1998, p. xii), but as valid and natural ways for students to demonstrate how they learn, and what they know and understand (Blecher & Jaffee, 1998). Our experiences show us that integrating literacy lessons with the arts helps meet the individual learning needs of all students and empowers those who come from diverse cultural, linguistic, and ethnic backgrounds. Weaving literacy instruction with arts activities promotes students' interests in reading and writing (McMaster, 1998) and provides opportunities for those with varied learning styles and special learning requirements to portray their emotions, explore and enhance their understandings, and extend and express their thoughts effectively.

Especially relevant to our work as professors of literacy, we note that aesthetic engagements create possibilities for students from nonmainstream backgrounds to construct, extend, and share meaning from the texts they read and to generate new meaning and novel texts (Graves, Juel, & Graves, 1998) (cf. Prest & Prest, 1988; Purves, 1993; and Rosenblatt, 1995, for discussions of transactional literary theory and readers' aesthetic responses to literature). In addition, participating in the arts gives students who speak variations of standard English opportunities to increase their oral and written language. They play with words, invent language, and create music from phrases (Steinbergh, 1995). Equally important, the arts provide cultural opportunities that are not related to social class, academic ranking, or standard English proficiency (Pankratz & Mulcahy, 1989).

Further Solidifying Our Convictions

Over the past few years, we have further solidified our convictions about the importance of supporting students' literacy learning with the arts. Insights from educational scholars have convinced us that as we enter the twenty-first century, language arts instruction must broaden "to reflect the greater oral, written, and communication needs [of students]" (Tompkins, 1998, p. 23). These increased communication requirements include the ability to comprehend and integrate visual material, such as films, photographs, commercials, and videos (i.e., viewing), with written forms of text, and the ability to create meaning through drama productions, murals, diagrams, and illustrations (i.e., visually representing) (Alvermann, Moon, & Hagood, 1999; Cairney, 1997; Flood, Lapp, & Wood, 1998; Tompkins, 1998). The concept of multiliteracy (Eisner 1992, 1997) and the current focus on multiple intelligences (Gardner 1991, 1993) also have redefined and extended our definition of literacy beyond print and oral language to encompass

drama, poetry, music, dance, and the visual arts as legitimate venues and expressions of understanding and communication (Blecher & Jaffee, 1998; Eisner, 1997). Thus, the arts occupy a central place in forming and extending all aspects of our curriculum (see Gallas, 1995).

Preservice Teachers Integrating Literacy Instruction with the Arts

Because of our firm beliefs about the benefits of arts encounters for urban, elementary students, as part of course requirements, preservice teachers in our reading/language arts field programs integrate literacy instruction with the arts at every opportunity. Following ideas from Vygotsky's "Zone of Proximal Development" (see Moll, 1993 and Vygotsky, 1978), the preservice teachers collaborate with their students in presenting student-authored puppet shows, Readers Theatre, and drama productions. They work side-by-side with their students, scaffolding, modeling, and creating text-based murals, story quilts, and songs, and as co-constructors of knowledge, along with their students, they author and publish fiction and informational texts.

Preservice Teachers' Cases

Our preservice teachers also write a teaching case describing their reflective thinking, hopes, concerns, problems, and achievements as they address the complexities of devising and offering literacy-based arts lessons (see Merseth, 1991; Richards & Gipe, 2000; J. Shulman, 1992; L. Shulman, 1992, for the benefits of case writing for education majors). Each two-to-three page narrative is written in the first person, contains dialogue, and presents relevant, context-specific details. We also urge our preservice teachers to provide closure to their cases by including some possible solutions to their instructional dilemmas and some explanations for their accomplishments (see Richards & Gipe, 2000 for ideas to help guide preservice teachers' case writing efforts).

Recently, we have become particularly drawn to the content of our preservice teachers' cases as rich sources of information. Specifically, we noted that the narratives provide a window into our preservice teachers' planning, thinking, and lived experiences as they undergird and nurture their students' literacy learning through the arts. We also recognized that when viewed as a collection of authentic teaching stories, the cases offer teachers and teacher educators considerable insights about the realities of linking students' literacy instruction with aesthetic encounters. Most importantly, we discerned that the narratives are grounded in veracity and credibility -- that is, they portray real preservice teachers who offer genuine literacy-based arts lessons in bona fide, urban elementary schools. Since there is a paucity of research concerning the integration of literacy education with the arts and arts

education in general (Eisner, 1998; Louisiana Institute for Arts in Education, 1998; Pankratz & Mulcahy, 1989; Taylor, 1996), thinking that we might make a contribution to the literature, we decided to conduct a systematic research project documenting the themes in our preservice teachers' narratives. We also hoped to expand our understandings about our two nontraditional field programs in which the curricula extend beyond conventional approaches to teaching reading and writing. Ultimately, we hoped to improve our own pedagogy by uncovering the concerns, dilemmas, needs, and triumphs of preservice teachers who work to develop urban students' literacy through artistic pursuits.

Conceptual Framework for Our Inquiry

Three literatures informed our inquiry: a) tenets of social constructivism which suggest that language reveals individuals' knowledge, perceptions, and beliefs (Goetz & LeCompte, 1984); b) ideas from discourse analysis that describe written texts as true reflections of human experiences (Gee, Michaels, & O'Connor, 1992); and c) premises from social interactionism which suggest that as mature human beings encounter problems that emerge through their circumstances, they move to resolve those problems through thoughtful reflection and action (Woods, 1992). In addition, we were mindful of traditions from hermeneutics which "indicate that the same text can be read [and interpreted] in a number of different ways (Tappan & Brown, 1992, p. 186).

Research Methodology

Questions Guiding Our Research As They Related to the Integration of The Arts with Literacy Instruction

In our inquiry we sought to answer the following questions:

- 1) What universal themes are visible in our preservice teachers' cases?
- 2) Are possible theme variations in our preservice teachers' cases related to the contextual conditions of the two different urban schools in which they work?
- 3) What are the most common problematic issues identified by our preservice teachers?
- 4) What teaching/learning successes and accomplishments are portrayed in the cases?
- 5) Do our preservice teachers' quandaries or accomplishments reside in distinct domains of the arts, such as drama, music, creative writing, or the visual arts?
- 6) Do the contents of our preservice teachers' narratives illuminate instructional shortcomings or program issues that we, as supervisors, need to remedy?
- 7) Are the implications of the research germane to teacher education and elementary school curricula?

Analyzing the Cases

As described earlier, each preservice teacher wrote a teaching case expressing his/her dilemmas interfacing literacy lessons with the arts. The preservice teachers worked in grades K-6. All came from middle-socioeconomic backgrounds; six were male and 197 were female. Their ages ranged from 21-45 years; eleven were African-American and 192 were Caucasian.

Working as a research team, we examined and categorized the 203 teaching cases written by the eight groups of preservice teachers who matriculated through our programs each fall and spring semester over the past two years. First, we collated the cases according to the two different teaching contexts in which our preservice teachers worked (97 cases were written by four groups of preservice teachers in Davis School, and 106 cases were written by four groups of preservice teachers in Logan School).

In subsequent meetings, using analytic induction (Bogdan & Biklin, 1992), we read and reread the narratives, looking for emerging categories and patterns that would "facilitate a coherent synthesis of the data" (Gay, 1996, p. 227). We made notes and underlined what we considered to be salient dimensions of the texts as a way of revealing the predominant theme or central issue contained in each case (e.g., "drama and management"; "art and reading lessons").

Next, we coded and categorized each narrative according to the prevailing theme (e.g., "Classroom teacher's apathy and disinterest in the arts"). Finally, we tabulated the number of cases in each thematic category for each school context. We settled any differences of interpretations through collegial discussions until we reached an agreement.

Themes in Our Preservice Teachers' Cases

Analysis of the cases revealed that across the two teaching contexts, our preservice teachers wrote about 11 major issues: a) problems caused by preservice teachers who were overly 'teacher-directed' in the arts; b) tensions related to classroom teachers' apathy and disinterest in the arts; c) concerns about students who created books and fictional pieces reflecting personal experiences with violence and stressful environments; d) dilemmas caused by preservice teachers neglecting to model and scaffold arts lessons for students; e) quandaries about students who were afraid to take risks in arts activities; f) successes concerning the increased motivation of disruptive or academically at-risk students when they became fully engaged in arts experiences; g) dilemmas about students' lack of familiarity with visual arts media; h) frustrations about older students' arts initiatives being significantly

influenced by popular, animated, high tech films and provocative rock and 'hip-hop' videos; i) dilemmas caused by students' lack of cooperation; j) quandaries associated with preservice teachers' inadequate planning and insufficient group management skills; and k) problems caused by preservice teachers offering arts activities that were not linked to literacy lessons (see Table 1 for the number of cases in each thematic category for each school context). The following case excerpts illustrate these common themes.

Table 1
Predominant Case Themes and Number of Cases
in Each Category for Each School Context

Themes	Davis School - 97 Cases		Logan School - 106 Cases	
	Davis School	n (%)	Logan School	n (%)
preservice teachers overly teacher-directed	10	(10.3%)	9	(8.5%)
classroom teachers' apathy	2	(2.1%)	1	(1%)
books/fiction reflecting students' stressful environments	8	(8.3%)	9	(8.5%)
preservice teachers neglecting to model/scaffold arts lessons	9	(9.3%)	10	(9.5%)
students afraid to take risks in arts activities	15	(15.5%)	14	(13.2%)
positive behavior changes in disruptive/at-risk students	6	(6.2%)	9	(8.5%)
students' lack of familiarity with visual arts media	10	(10.3%)	8	(7.6%)
older students' art initiatives influenced by film and 'hip-hop' videos	4	(4.2%)	6	(5.7%)
students' lack of cooperation	9	(9.3%)	13	(12.3%)
preservice teachers' inadequate planning/management skills	21	(21.7%)	23	(21.7%)
preservice teachers offering arts activities unrelated to literacy lessons	3	(3.1%)	4	(3.8%)

Problems Caused by Preservice Teachers Who Were Overly 'Teacher-Directed' in the Arts

Drama Needs Democracy

When I first told my students that we would be performing a play, they immediately became disruptive and started yelling out what play they wanted to perform and what parts they wanted to take. Well, I got really upset over their behavior so I said, "The next time we meet I will have our drama production typed and I will assign parts to everyone."

At our next session, I announced, "We are going to do a play about *Goldilocks and the Three Bears*. Jonah,

you are the Poppa Bear, Margie, you are the Momma Bear, and George, you are the Baby Bear. Salina is Goldilocks and Mercedes is the narrator."

I thought that my prior planning and concrete directions would solve everything, but was I wrong. Jonah said, "I want to be the narrator," and Salina said, "I want to be a bear."

So, mass confusion reigned once more. Then, when we got the parts straightened out, the group continued to be divided about their lines, and they argued about where they should stand on the stage. Their disruptive behavior continued throughout all of our play practices. What a mess! I finally went to the library and read *Start with a story* (Watson-Ellam, 1991). I learned that creating drama productions should bring about a sense of community among participants rather than chaos. I also found out that I shouldn't tell my students what play we will present or assign parts to them. Rather, I need to discuss drama possibilities with my students, serve as a resource as they create their drama productions, and then, give them choices about what characters they will play.

Tensions Related to Classroom Teachers' Apathy and Disinterest in the Arts

Teachers Not Getting Involved

While my group was working on our mural, the classroom teacher and her assistant were gossiping, drinking coffee, and eating girl scout cookies. How inappropriate is that? After our mural was finished I asked the teacher to look at our work. I was really shocked when she pointed to Mariel's starfish and Noah's lobster and said, "Now what's that supposed to be?"

Then, she went back to drinking coffee. I will not let this teacher's lack of enthusiasm for our art work affect my thinking about teachers, teaching, and literacy-based arts lessons. I spoke to a teacher who has been teaching for more than twenty years, and she said to just remember the good things about this project and never act like that classroom teacher.

Concerns about Students Who Created Books and Fictional Pieces Reflecting Personal Experiences with Violence and Stressful Environments

What You See Is What You Talk about And What You Dream at Night

First we read a great book . . . *Who lost a shoe* (Hazen, 1992). Then we talked about the important parts of stories. The kindergartners really grasped the concept of characters, settings, problems, and solutions so I decided they were ready to dictate their stories to me. Roger was first. He said, "Somebody stold my shoes and

"FIRST WE READ A GREAT BOOK:"

I asked my momma for new ones and she said, 'No,' so I went to steal some shoes."

Next was Tyrone. He said, "My sister peed in my momma's bed and my momma whooped her until she cried and fell asleep and didn't wake up for a long time."

Devonte told his story by saying, "They crashed in the windows of my grandmaw's truck and scart me in the middle of the night. My brother was running out and yelling. Then, they stold the truck."

I sat very still for a minute. I did not want to continue with these stories. I was distraught. I didn't know what to do, so I said, "We'll finish these stories the next time we meet."

None of my fellow preservice teachers knew what to do about these poor children so I went to the library and skimmed a book . . . *Violence - Our fastest growing public health problem* (Langone, 1984). This book shares the opinion that what you see and experience in your environment is what you talk about during the day and what you dream about at night. When the children told me stories about violence, they were really telling me about their everyday realities. How can I help these poor children?

Dilemmas Caused by Preservice Teachers Neglecting to Model and Scaffold Arts Lessons

I Knew Better/Why Didn't I Model?

When it was time for my fourth grade students to write their books, it never dawned on me that they did not understand the basic elements of a story. I explained, "We will be writing and publishing our own books." Then, I asked, "Who will the authors of our books be?"

Instead of a happy response all I got was eight kids with blank stares. I had planned for our lesson to be a prewriting session, and I suddenly realized that my plans were going to have to go out the window because none of my fourth graders knew anything about making books. What was I thinking? All I have heard from my professor is to "Model, model, model." And . . . I didn't model anything. Naturally, this lesson was awful. Well, I learned a lesson myself . . . never take anything for granted about students' background knowledge and never forget to MODEL!

Quandaries about Students Who Are Afraid To Take Risks in Arts Activities

Fear of Failure

My fourth graders recently constructed a mural based upon the story *Charlotte's Web* (White, 1952). All of the students did a lovely job except Michael. He is an excellent student who has to be perfect in everything. In

fact, he can't do any visual arts activities unless he has a ruler in his hand. He is so uptight about making a mistake that he detests art. For example, when we were doing the border for our mural, Michael stood back and let others in the group paint. I asked him, "Michael, don't you want to join in?"

Michael shook his head and replied, "Nope, I might mess up. I might make the border crooked."

Michael also is insecure about writing creatively. Last week I reviewed the basic features of stories and showed examples of other students' creative books. Then, I explained to my group that we would write and publish our own books. Michael immediately began to complain. "I don't know what to write about. I can't draw," he said.

No matter how I encouraged him, he refused to paint, write, or draw. I asked my professor about Michael, and she said that some students have few opportunities to work with visual arts media. Therefore, they don't know how to paint or draw. She also said that students may not enjoy writing because they have a fear of failure or they lack personal awareness of what steps to take to write. I know Michael is afraid to fail in the arts, but I don't know what to do about it.

Successes Concerning the Increased Motivation of Disruptive or Academically At-Risk Students When They Became Fully Engaged in Arts Experiences

Success At Last

Poor Carlos had problems with every literacy activity I introduced this semester. He is a third grader who can hardly read and write. As the semester progressed, I could see Carlos' spirit decreasing, and I felt like I was letting him down. I teamed him up with a learning buddy, but it seemed like it was too late. Carlos gave up trying and he started to be a behavior problem. At the end of March we began working on our drama production. Well, who would have thought that the arts could have such a positive effect on a student?

First, we picked names to assign characters. When it was Carlos' turn he excitedly said, "I want to be the big, bad wolf!"

As the weeks went by, we practiced over and over. One day, Carlos stopped me in the hall. He said, "Miss Clarke, I've been thinking about my part as the wolf. Can I change some of my lines in the play so they will be easier for me to remember?"

"Sure," I replied.

The following Tuesday, I went into the auditorium to get my group ready to perform and there was Carlos in his wolf costume with the biggest grin I ever saw. I was not used to seeing him smile and it was a shock! Carlos had "confidence" written all over his face. Well, we got up on

the stage and he made me so proud! It was as if I was watching my own child perform. After our final bows, I gave him a big hug and then it hit me. Performing in a drama activity had provided Carlos with the first school success he's had since I've known him. I bet Carlos could learn to read and write by authoring, producing, and starring in plays.

Dilemmas about Students' Lack of Familiarity with Visual Arts Media

The Great Pyramid Mess!!

My fifth grade students were all painting three-dimensional pyramids described in the book *Emily Eyefinger and the lost treasure* (Ball, 1994). They loved painting and were really involved in their work. However, I didn't pay enough attention to what they were doing. When I glanced over at their work, they had mixed all the paint together until they formed a greyish-green color instead of the red and yellow we needed for our pyramids. Well, they certainly had a good time exploring how to mix and use paint, but I had to throw out all the paint and begin the lesson all over again.

I found some solutions to this pyramid/paint mess by reading *The effective teaching of language arts* (Norton, 1997). In this text, a man named Ernst (1994) describes an artists' workshop where students look at the work of other artists before they begin painting and drawing. This pre-drawing and painting activity helps students become familiar with various types of media. Ernst also suggests that teachers provide mini-lessons that focus on techniques of painting and drawing before students engage in these activities. When students have opportunities to manipulate arts media, they learn how the media works. If I had presented some pre-art lessons, we wouldn't have had such a mess.

Frustration about Older Students' Arts Initiatives Being Significantly Influenced by High Tech Films and Provocative Rock and "Hip-Hop" Videos

'Hip-Hop' and All That Rock

It isn't easy offering literacy-based arts activities to older students who have had few opportunities to engage in artistic pursuits. I teach sixth grade boys and most have never made books, pasted, sung in a school chorus, painted, created papier mâché sculpture, or participated in drama productions. They love our literacy-based arts experiences, but they have no idea how to use their own self-expression and creativity. Instead, they fall back on what they see on TV and hear on their boom boxes.

"Dr. Love and Puff Daddy are COOOL," David shouted when we were talking about putting on a play.

"Yeah," Anthony responded, "Let's do a play about Dr. Love and Puff Daddy and their wicked ways."

"I don't know who they are," I answered. The idea of putting on a play about rap stars named Dr. Love and Puff Daddy were not on my teaching agenda.

The same thing happened when we painted. The students created terrible looking creatures that they had seen on TV. They also tried to paint really provocative looking women that looked like the Spice Girls, and they got frustrated when they couldn't create animated cars, trains, and animals like they see in high tech films. Well, I finally had to make some rules. "No more TV and rap stuff in our work," I announced. "We need to use our own imaginations and ideas."

I wonder if I am fooling myself? These students are significantly influenced by television, high tech films, and 'hip hop' and rock music as Alvermann, Moon, and Haygood state in their text *Popular culture in the classroom: Teaching and researching critical media literacy* (1999). How can I counteract what they see and hear daily and consider "cool"? How can I bring out their own creativity? I can't find any solutions to this problem at all.

Dilemmas Caused by Students' Lack of Cooperation and Inappropriate Behavior

Why Were They So Mean to Jim?

We were working on our mural, and every one of my third graders was painting the background portion as a team. The next task was to decide which students would paint each of the story characters. This part went well too. I remember thinking, "Wow, things are going so smoothly."

All of a sudden, things turned ugly! Jim, was working diligently on his illustration and Cedric turned to him. "EEWW!!" he said, "What are you making?"

Poor Jim felt terrible . . . I could tell by the expression on his face.

To make matters worse, some of the other students joined in. "Eew Cedric . . . You right. We ain't gonna put Jim's stuff on our mural cause it will mess up the whole thing."

Of course, I stuck up for Jim. I said, "Jim, your art work is fine." I also told the other students, "You are being mean and uncooperative."

This event remained in my brain all week. I kept questioning myself about why the students had ganged up on Jim. Should I have handled this situation differently? Was there more I should have done? After reading a book about the personality dimensions of eight-to-ten-year-olds, I learned that they are very aware of others' mistakes and they are not very sensitive to others' feelings. I also learned that students this age display a robust capacity to

heal quickly. Sure enough, at our next meeting, the students were all getting along. But, I wonder how I could have handled the students' uncooperative and inappropriate behavior better.

Quandaries Associated with Preservice Teachers'
Inadequate Planning and Insufficient
Group Management Skills

What Did We Do Wrong?

For our drama presentation we combined all three of our kindergarten groups together. The song we chose to include in our play had repetitive lyrics, and we thought it would be easy for our kindergarten students to learn. We practiced that song day after day for thirty minutes at each session. But, each time we practiced, our management problems got worse. "I want to be in the front of the bus," Patrice would always yell.

Tyrone always complained, "He pushed me."

Other children ran up and down the stage and some even jumped off the edge of the stage to the floor below.

It went on and on like that until finally we decided that whoever spoke or acted out of turn would be put out of the group. We actually ended up with only 12 children by the last practice. Of course when it was the day of our presentation we relented and let all of the children participate. I know we didn't do a good job as preservice teachers and co-producers of a play. A dancing teacher told me that rehearsals should never be longer than fifteen minutes or children lose interest. Maybe that's one reason that we couldn't keep order. But, lurking in my mind is another reason for our dilemma. I think we didn't take charge of our large group.

Problems Caused by Preservice Teachers
Offering Arts Activities That Were
Unrelated to Literacy Lessons

Is It Time For Art Yet?

After about thirty minutes into every lesson my third graders always asked, "Miss Jennifer, is it time for art yet? Are we going to have art?"

I always said "Yes, look at our 'News of the Day.' Art is listed on our chart."

Then, worried that my students were losing interest in our reading lesson, I always asked them, "Don't you want to finish reading our story and doing our writing activity?"

Usually, the students would agree to finish the reading lesson, but I always sensed that they saw reading and writing only as a way to get to our art activities . . . in other words, reading and writing were just a means to an end for them . . . art was really what they wanted to do. In

fact, when they were involved in the arts, they always calmed down and concentrated the most.

During one lesson, after reading a collection of Shel Silverstein's poetry, I asked my students to illustrate their favorite poem. As I walked around, interacting with my group, I noticed that Samantha was drawing a really fine picture of a little girl. "What a beautiful picture," I said, "Who is the little girl?"

Samantha proudly replied, "It's my sister."

"Your sister? What poem is that from?" I asked.

With a confused look on her face, Samantha said, "It's not from a poem. It's my sister."

I moved on and noticed that Rachel was filling her paper with hearts, flowers, and sentences that said, "I love you."

"Is this from a poem we read?" I asked her.

"Oh no," Rachel replied. "I'm making this for you."

I tried to explain to the group what we were supposed to be doing. But, Rachel thought I didn't like her picture and she began crying.

I continued looking at my students' work and discovered that every single student was drawing whatever they felt like drawing. I suddenly realized that they had been drawing and painting whatever they liked the entire semester. They hadn't a clue that their art work should be connected to our literacy lessons. They viewed art as "free time." I spoke to my professor about this dilemma and she said to make sure that my students always know why they are engaged in arts activities. I also recognized that I had contributed to my students considering art as "free time" by saying in each lesson, "Hurry through your writing so we can get to art."

This reinforced my students' perceptions that our art work was divorced from our literacy activities. I also caused problems by always handing out art supplies prior to giving directions. By the time I gave directions, my students were already drawing and painting and they didn't listen to me. Further, I never engaged my students in productive discussions that would help them link our work in literacy with the arts (e.g., an exchange of ideas about their favorite Shel Silverstein poems).

Discussion

Explorations of our preservice teachers' cases illuminated 11 prevailing themes. Ten themes related to problematic issues confronting our preservice teachers as they worked to support their students' literacy learning through the arts. A single theme highlighted the successes and joys of our preservice teachers as they recognized the increased motivation and enthusiasm of disruptive or academically at-risk students when they became fully engaged in aesthetic activities.

Our research showed that across the two teaching contexts (i.e., Davis and Logan Schools), these 11 themes were uniformly distributed throughout all of the arts domains (i.e., writing, and the visual and performing arts). Therefore, in this study, distinctions in school setting did not produce significant theme variations in our preservice teachers' cases (see Table 1 for the number of cases in each of the theme categories for each school context).

One explanation for the stability of the themes across teaching contexts and arts domains is that the structure, course requirements, and focus of our two field initiatives are comparable. In addition, as supervisors, our expectations for our preservice teachers' achievements are similar and our teaching philosophies and instructional styles are parallel. Another consideration is that although Davis and Logan Schools are located in different states, their instructional policies are equivalent. For example, in both schools, few arts activities are offered, and students are expected to work quietly and individually. Group collaborations and interactive discussions are discouraged. Further, the majority of students in Davis and Logan Schools live in harsh neighborhood environments. Violence and poverty are common and students' life circumstances manifest themselves through their conversations, stories, behavior, and artistic expressions.

The stability of case themes also may relate to the neophyte status of our preservice teachers. In both schools a preponderance of preservice teachers wrote about group management issues. Research shows that a prevalent concern for beginning teachers is how to understand and guide unmotivated pupils and how to manage groups of students (Richards & Gipe, in press; Robeck & Wallace, 1990).

It is not surprising that a second dominant theme in our preservice teachers' cases centered around students' apprehensions about taking risks in creative activities. Art is basically excluded from the regular curricula in Davis and Logan Schools, and as a result, students have little knowledge about how to use visual arts materials, perform on the stage, compose songs, or write creatively in order to express their personal thoughts, understandings, and emotions. Students who perceive themselves as inadequately prepared in the arts would naturally feel reticent and cautious about engaging in artistic pursuits.

Implications for Teacher Education and Elementary School Curricula

The findings of our inquiry attest to the efficacy of case writing for preservice teachers. Concentrating on educational quandaries and pondering possible solutions to problems has the capacity to facilitate preservice teachers' professional growth. (L. Shulman, 1992). Further, the themes in preservice teachers' cases can serve

as windows into their teaching experiences. For example, we learned what dilemmas confronted our preservice teachers as they supported their students' literacy learning through the arts. In addition, we now have narrative evidence that artistic pursuits can benefit disruptive or academically at-risk students.

The question then becomes, what do we, as literacy teacher educators, need to do to help our preservice teachers effectively integrate students' literacy learning with the arts? Careful attention to the case issues that preservice teachers consider important certainly provides insights about our own competence in guiding preservice teachers' professional development. Through our case analysis, we noticed gaps in our course instruction. In particular, we learned that in our lectures, we need to stress the importance of modeling literacy-based arts lessons, and that it is crucial for us to demonstrate these types of integrated lessons for our preservice teachers. We also must supply relevant readings and seminar discussions with appropriate guest speakers to help our preservice teachers understand and appreciate the linkage between their students' backgrounds, interests, and artistic expressions. Equally important, we need to increase our demonstration lessons for our preservice teachers to help them become more accomplished in managing groups of students, more proficient in connecting literacy lessons with relevant arts activities, and more responsive and able to empower students who are afraid to take risks in the arts. Further, we must consider the feasibility of adding arts-based workshop sessions or cross-disciplinary seminars to our course agendas that provide our preservice teachers with practical experiences working with various types of visual arts media, authoring fictional and information books, and participating in informal drama activities. Hands-on encounters with the arts have the potential to awaken preservice teachers' aesthetic perspectives and help them become more responsive to their students' artistic needs and interests.

The research also underscores the potential dangers of reducing or eliminating the arts in elementary education. Instead of "widening the learning circle" (Blecher & Jaffee, 1998, p. xi), we run the risk of disallowing many students full engagement in learning and the curriculum. As Cairney (1998), Eisner (1997), and Greene (1995), point out, there are many forms of literacy. In the twenty-first century literate students will not simply read and write (Cairney, 1998). Literate students will demonstrate confidence and competence in using all types of media as essential forms of communication, including technology, print, and the visual and performing arts. Schools that deny students access to the arts will limit the ways in which students can access meaning, learn, think, and communicate.

References

- Alvermann, D., Moon, J., & Haygood, M. (1999). *Popular culture in the classroom: Teaching and researching critical mass media*. Newark, DE: A joint publication of the International Reading Association and the National Reading Conference.
- Ball, D. (1994). *Emily Eyefinger and the lost treasure*. New York: Simon and Schuster for Young Readers.
- Blecher, S., & Jaffee, K. (1998). *Weaving in the arts: Widening the learning circle*. Portsmouth, NH: Heinemann.
- Bogdan, R., & Biklin, S. (1992). *Qualitative research for education (3rd ed.)*. Needham Heights, MA: Allyn and Bacon.
- Bolton, G. (1985). Changes in thinking about drama in education. *Theory into Practice*, 24, 151-157.
- Cairney, T. (1997, September). New avenues to literacy. *Educational Leadership*, 76-77.
- Eisner, E. (1992, April). The misunderstood role of the arts in human development. *Phi Delta Kappan*, 591-595.
- Eisner, E. (1997, January). Cognition and representation: A way to pursue the American dream? *Phi Delta Kappan*, 349-353.
- Eisner, E. (1998, January). Does experience in the arts boost academic achievement? *Art Education*, 51(1), 7-15.
- Ernst, K. (1994, January). Writing pictures, painting words: Writing in an artists' workshop. *Language Arts*, 71, 44-52.
- Flood, J., Lapp, D., & Wood, K. (1998, November). Viewing: The neglected communication process or "When what you see isn't what you get". *The Reading Teacher*, 52(3), 300-304.
- Fowler, C. (1988). *Can we rescue the arts for America's children?* New York: American Council for the Arts.
- Gallas, K. (1991, February). Arts as epistemology: Enabling children to know what they know. *Harvard Educational Review*, 61(1), 40-51.
- Gallas, K. (1995). *Talking their way into science: Hearing children's questions and theories, responding with curricula*. New York: Teachers College Press.
- Gardner, H. (1991). *The unschooled mind: How children think and schools should teach*. New York: Basic Books.
- Gardner, H. (1993). *Creating minds: An anatomy of creativity seen through the lives of Freud, Einstein, Picasso, Stravinsky, Eliot, Graham, and Gandhi*. New York: Basic Books.
- Gay, L. (1996). *Educational research: Competencies for analysis and application (5th ed.)*. Englewood Cliffs, NJ: Prentice Hall.
- Gee, J., Michaels, S., & O'Connor, C. (1992). Discourse analysis. In M. LeCompte, W. Millroy, and J. Preissle (Eds.). *The handbook of qualitative research in education* (pp. 227-291). New York: Academic Press.
- Goals 2000: The Educate America Act (1994). Report on H R. 6, Committee on Educational and Labor, House of Representatives, 103rd Congress, 2nd Session (ERIC Document No. ED 369 147)
- Goetz, J., & LeCompte, M. (1984). *Ethnography and qualitative design in educational research*. Orlando, FL: Academic Press.
- Graves, M., Juel, C., & Graves, B. (1998). *Teaching reading in the 21st century*. Boston: Allyn and Bacon.
- Greene, M. (1995). *Releasing the imagination series. Essays on education, the arts, and social change*. San Francisco: Jossey-Bass.
- Hazen, B. (1992). *Who lost a shoe?* New York: Newbridge Communications.
- Langone, J. (1984). *Violence!: Our fastest growing public health problem*. Boston: Little Brown.
- Louisiana Institute for the Arts in Education. (Personal Correspondence, 1998). 225 Baronne Street, New Orleans, LA 70112.
- McMaster, J. (1998, April). "Doing" literature: Using drama to build literacy classrooms: The segue for a few struggling readers. *The Reading Teacher*, 574-584.
- Merseth, K. (1991, September/October). The early history of case-based instruction: Insights for teacher education. *Journal of Teacher Education*, 42(4), 243-249.
- Moll, L. (1993). *Vygotsky and education: Instructional implications and applications of sociohistorical psychology*. New York: Cambridge University Press.
- Norton, D. (1997). *The effective teaching of language arts (5th ed.)*. Columbus, OH: Merrill Prentice Hall.
- Pankratz, D., & Mulcahy, K. (Eds.). (1989). *The challenge to reform arts education: What role can research play?* ACA Arts Research Seminar Series. New York: American Council for the Arts.
- Popp, M. (1996). *Teaching language and literature in elementary classrooms*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Prest, P., & Prest, J. (1998). Theory into practice: Clarifying our intentions: Some thought on the application of Rosenblatt's transactional theory of reading in the classroom. *English Quarterly*, 21(2), 127-133 (ERIC Document No. ED 378 672)
- Purves, A. (1993). *Towards a reevaluation of reader response to school literature*. New York: National Research Center on Literature Teaching and Learning (ERIC Document No. ED 395 515)

- Richards, J. & Gipe, J. (2000). *Elementary literacy lessons: Cases and commentaries from the field*. Mahwah, NJ: Lawrence Erlbaum, Associates.
- Riley, R. *Mississippi Sun Herald, Gulfport, MS*, (November 11, 1998, Section B, p. 1).
- Riley, R. *Mississippi Sun Herald, Gulfport, MS*, (November 12, 1998, Section B, p. 4).
- Robeck, M., & Wallace, R. (1990). *The psychology of reading: An interdisciplinary approach* (2nd ed. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Rosenblatt, L. (1995). *Literature as exploration* (5th ed.). New York: Modern Language Association.
- Shulman, J. (Ed.). (1992). *Case methods in teacher education*. New York: Teachers College Press.
- Shulman, L. (1992). Toward a pedagogy of cases. In Shulman, J. (Ed.). (1992). *Case methods in teacher education* (pp. 1-30). New York: Teachers College Press.
- Steinbergh, J. (1995). To arrive in another world: Poetry, language development, and culture. In M. Goldberg and A. Phillips (Eds.). *Arts as education* (pp. 33-53). Cambridge MA: Harvard Educational Review.
- Tappan, M., & Brown, L. (1992). Stories told and lessons learned: Toward a narrative approach to moral development and moral education. In C. Witherell and N. Noddings (Eds.). *Narrative and dialogue in education* (pp. 171-192). New York: Teachers College, Columbia University.
- Taylor, P. (Ed.). (1996). *Researching drama and arts education: Paradigms and possibilities*. London: The Falmer Press.
- Tompkins, G. (1998). *Literacy for the twenty-first century: A balanced approach*. Upper Saddle River NJ: Merrill Prentice Hall.
- Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Watson-Ellam, L. (1991). *Start with a story: Literature and learning in your classroom*. Portsmouth, NH: Heinemann.
- White, E. B. (1952). *Charlotte's web*. New York: Harper Trophy.
- Woods, P. (1992). Symbolic interactionism: Theory and method. In M. LeCompte, W. Millroy, and J. Preissle (Eds.). *The handbook of qualitative research in education* (pp. 337-404). New York: Academic Press.

A Critique of Methods Used to Answer the Question: Does Money Make a Difference?

Jim C. Fortune and Susan Spofford-Richardson
Virginia Tech

The relationship between per pupil expenditures and pupil achievement is an interesting and important policy consideration. Six methods have been used to address this issue. These methods have produced inconsistent results because they violate traditional research methodology. Similarly, two different methods used to summarize the results have produced conflicting conclusions. These research methods and summaries are critiqued and the violations of research assumptions and methods are explained. Results of this critique suggest a preponderance of evidence supporting the existence of a positive relationship between expenditures and achievement.

School equity suits or budget disputes almost always cause someone to ask the question "Does money really make a difference in school achievement?" On the surface, this question may seem too difficult to answer. However, the answer to the question is usually not as helpful as anticipated and is much harder to develop than one might expect. Five methods (please refer to page 1, column 2) have traditionally been used to address this question. These techniques emerged since the large regression analysis of Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, & York (1966) produced a basis for federal intervention in education and Hanushek (1986, 1989) decided in the late 1980s to popularize the notion that there is too much waste in the public schools for money to matter.

This type of search for policy answers is generally made futile by a failure to appreciate the nature of statistical inference and an absence of policy parameters required to distinguish trivial from meaningful relationships. Investigators are generally forced to look for a statistically significant effect within a post hoc design, assuming that the size of an effect is relevant without input from policy makers. If an effect is found, the results are subject to debate regarding their utility. If no such effect is found, researchers tend to argue that no relationship exists between money and achievement. Our understanding of statistical inference suggests the

null hypothesis, that no relationship exists between money and achievement, is assumed to be true. Finding an effect permits rejection of this hypothesis, but not finding an effect does not permit its acceptance. In nonstatistical terms, there are many reasons to reject a null hypothesis. Among such reasons are small sample size, lack of measurement precision and variable specification, and a failure to control intervening variables. If sample size is very large or if measurement is overspecified, then observed effects may in fact be trivial. Statistical tests alone are insufficient for determining policy relevance.

Five research methods and two summary methods have been used to address the question, with different levels of success and with different conclusions. These research methods include: (a) the linear correlation method of production function analysis; (b) the multiple regression method of production analysis; (c) a global across time comparison model; (d) a threshold driven comparison model; and (e) a structural equation model. The summary methods include the counting method and the meta-analysis method of literature summary.

Research Assumptions and Issues Pertinent to All Methods

Four research questions that result from unexplored assumptions need to be investigated. These questions are as follows: "Are the conditions assumed by the researcher(s) for the types of analyses conducted satisfied by the data?" "Are the data studied independent of confounding research conditions?" "What effects does data aggregation have on the outcomes?" "Do disparities in averages reflect true disparities?"

One deficiency in the application of all of these methods appears to be laxity in assuming that the condition being analyzed actually exists in the subject being studied. For example, student achievement is analyzed in

Jimmie C. Fortune is a Professor in the Department of Educational Leadership and Policy Studies at Virginia Polytechnic Institute and State University. Susan Spofford-Richardson is a doctoral candidate in the Department of Educational Leadership and Policy Studies, also at Virginia Polytechnic Institute and State University. Correspondence regarding this article should be directed to Susan Spofford-Richardson, 3348 Beech Tree Lane, Falls Church, VA 22042-4004 or via email to spofford@hotmail.com.

school districts with different levels of funding, to determine whether there are differences due to funding. However, none of these research models includes any attempts to assure that the students in the study sample experienced learning in the reporting district or under the conditions being studied. Differences in high school achievement may originate in elementary schools, which may not be included in the study.

None of the methods in question can address causation because all are based on *ex post facto* study. However, causation may not be needed for policy relevance. In a state of nature one must assume that all school districts have an equal opportunity for funding and that any observed inequality of funding must have occurred some way other than naturally. One or more points of research critique and logical scrutiny limit each of these methods. In many cases, studies were concluded because of special circumstances or problems that could influence the study's outcome. In other cases, results were used out of context and without adequate control. Both summary methods suffer from the fact that the studies used in the summaries range across grade levels, measures of school output, definitions of school resources or money, reasons for conducting the study, and levels of the independence in the findings extracted. Hanushek's (1981) summaries are more vulnerable to the last defect than those of Greenwald, Hedges, and Laine (1994, 1996). The latter pointed out that, "While we do not regard multiple analyses of data on the same individuals to be as informative as analyses of independent data sets, Hanushek treats the two as equivalent" (Greenwald et al., 1996, p. 411). Many of the studies included in both summaries were not conducted to assess the influence of funding on achievement. Rather, they were conducted for other purposes, which may affect the relevance of their results to address the policy question.

Most of the research studies addressed funding and achievement at a district or school level. However, achievement occurs at the student level, which must have already undergone one or two levels of data aggregation. One cannot know the effects of these data aggregations on the relationships of interest.

All of the studies defined funding disparities at the level of school districts. We cannot know whether there is a greater variance in the "within" level of expenditures than in the "between" level. Lack of uniformity in expenditures within districts may defeat the comparison of expenditures across districts.

Linear Correlation Method of Production Function Model

Application of the simple correlation model must meet the data assumptions required by the correlation

process and the limitations to inference assumed in the use of the model. Pearson's Product Moment Correlation requires that the observations be bivariate normal. The coefficient obtained is representative of the linear relationship between the two variables and measures association rather than causation. The simple linear correlation model was popularized in the early 1970s. Unlike the more complex models, the simple correlation model permits one to speculate about potentially beneficial policy decisions. One can easily see what size changes in output are expected for given changes in input. Many simple linear correlation production function analyses were taken from studies undertaken for reasons other than direct analysis of educational efficiency. Most of these studies have been included in the summaries developed by Hanushek (1981). Examples include the following: Beiker and Anshel (1973), who applied the production function model to rural schools through multiple correlation on individual students; Burkhead (1967), who looked at large city high schools; Cohn (1975), who looked at school districts within a state; and Raymond (1968), who looked at elementary and secondary schools in a state. Hanushek's (1981) summaries used simple correlation production functions taken from an array of studies including those just mentioned, the Equal Education Opportunities samples, and the Project Talent database. Many of these studies used regression analysis as the principal analytic tools, and reporting articles did not include first-order correlations, which leads one to assume that Hanushek applied the linear model to obtain the correlations reported in his summaries. He did use the direct correlation model and the policy expansion procedure in speculating about costs of achieving equity in his testimony in the Alabama Court Hearings on funding equity for the state (recorded in his deposition).

Inadequate Variable Specification

Perhaps the greatest problem in the use of the simple correlation model is variable specification. Problems in variable specification occur in the subsequent four ways: (a) failure to identify the components used to construct a given variable, (b) inclusion of confounding elements in the data, (c) attribution of characteristics not present in a variable, or (d) collections of one or more of the preceding problems.

Variable specification problems occur when it is not clear how a variable is constructed or what it contains. For example, the variable "per pupil expenditures" is created by a ratio. The source of the numerator and the denominator are important, and one must ask what items are included in the calculation of expenditures. Does this numerator include only expenditures for instruction or

DOES MONEY MAKE A DIFFERENCE?

does it include capital outlay, federal grant monies, or special program monies for untested students (as is the case of special education in some states)? What makes up the denominator? Does "enrollment" count average daily attendance, average daily membership, or weighted student enrollment?

Variable specification problems also occur when confounding data elements are included in the measurement of the variable. One example is the inclusion of special education students' test scores in the achievement test scores for a district when federal dollars for special education are removed from the expenditures. One state in which the authors completed analyses used four distinct methods to capture enrollment data. Per pupil expenditures could vary substantially, depending on the method used to capture the base enrollment data. Most studies do not make it clear what divisor was used in the calculation of this variable: average daily membership, average daily attendance, weighted daily attendance, or enrollment counts. It is also unclear what dollars were included in the numerator of this variable. Policy briefs should consider how the variables studied were constructed. Such information must be gathered from the original researchers, when they can be found.

The third variable specification problem occurs when characteristics are attributed to a variable but are not present in that variable. An example of this problem is the assumption that teachers who are more highly paid provide better teaching. Frequently, most school districts use a pay schedule based on seniority. However, teacher seniority does not automatically equate to more effective teaching techniques. Additionally, class size is another variable that could have hidden meaning. Small classes may exist to promote instruction to all students, but the classes also may serve a specialized few, such as those in need of special education, remedial, or bilingual instruction, or support for behavioral disorders. How the districts create these small classes is also important. If small classes are created by making other classes large, one can expect to see a different effect than if the small classes are created by increasing staff. Small classes created by making other classes large are not expected to result in changing mean achievement to the same degree as would small classes created by increasing staff.

Forced data aggregation occurs when the measurement of a variable is made at a level lower than the unit of analysis. District expenditures are frequently reported as a single aggregate per pupil expenditure. However, performance testing is done at the student level and must be averaged to create district-level data. Average values for a variable lose some descriptive value in that variances of these averages are not included in the analysis. Data

aggregation also tends to alter the distribution of variables that contain such aggregation. Aggregation of per pupil expenditure at the district level may create trivial differences between districts with regard to that variable. Also, aggregation and the use of a single per pupil expenditure measure may obscure differential spending between grade levels or at different schools.

What Constitutes Differences in Expenditures?

"Throwing a bucket of water on a raging fire will not keep a building from burning to the ground," but no one would argue on the basis of this experience that water has no value in fire fighting. The value of water is apparent only when enough is applied to overcome the fire. An analogous situation often occurs in education. Frequently, we judge an intervention strategy to be ineffective before we have really implemented a program that is intense enough to achieve the desired effects. "Compensatory education" is a case in point (Bridge, Judd, & Moock, 1979). The above phenomenon is called a threshold effect, which is one reason why the correlation method of production function analysis does not detect achievement differences due to small differences of funding. One dollar difference in funding will not purchase an observable difference in achievement. Instead, some collective difference in funding is required to purchase observable differences in achievement. With today's costs of education one may estimate the necessary increased amount to be greater than \$700.

Perhaps the greatest problem in the use of the simple, linear correlation method beyond variable specification is the absence of the cost disparities that are essential to demonstrate differences in educational purchasing power. An ordering of districts by the amount of instructional expenditures does not necessarily order districts by educational purchasing power. Ordering districts by dollar differences that are less than the measurement error associated with expenditures results in gross under-estimation of the true relationship between costs and achievement.

The Truncated Variable (Attenuation)

Truncation of variables can have important effects on study outcomes, particularly in its effect on dependent or outcome variables. Truncation occurs when the total range of a variable is not included in the analysis. Percent of students passing a given test, used as a measure of achievement, represents a somewhat unusual form of truncation. The variance in achievement is reduced greatly when percent passing is used instead of actual scores. The achievement measure is limited to a dichotomy (pass-fail) rather than the full variation of the actual scores. Variable truncation also occurs when the tests have either floor or ceiling effects. Ceiling effects occur

when the tests are too easy and a large percentage of participants achieve the maximum score. Floor effects occur when a large percentage of the total distribution score near zero on the test. These testing effects occur when specific subgroups (at risk students, for example) are used in the analysis.

The Multiple Regression Production Function Model

A second production function analysis model is based on regression procedures, where a single output variable is predicted by one or more input variables (usually expenditures, teacher experience, or other resources) and intervening variables (socioeconomic variables and school size, for example). This type of analysis uses the intervening variables to control for factors that may confound the true relationship between inputs and outputs. Some studies using multiple regression are directed toward finding sets of predictor variables for school effectiveness to be used later in school reform. The primary problems associated with this method along with variable specification are order of entry with shared variance and/or multicollinearity.

Coleman, who is often cited as the originator of production function analysis, used regression analysis as his basic analytic method. Other investigators who use this form of the production function include Boardman, Davis, and Sanday (1977); Brown and Saks (1975); Cohn (1975); Ferguson (1992); Tuckman (1971); Walberg (1983); and Winkler (1975). At least half of these studies address other research issues and do not utilize production function analysis as the major aspect of the study.

Order of Variable Entry and Shared Variance

The order of variable entry in a calculation is important in handling shared variance or commonality of explanation. The potential for shared variance problems exists whenever two independent variables are related to a dependent variable and to each other. In the methodology where ordering is used to control for intervening variables, the first variable to enter into the computation gets credit for the correlation to the dependent variable that it shares with the second variable, which is likely to be an independent variable.

In Walberg's (1983) New Jersey study, the shared variance problem confounds his analysis. If size and socioeconomic variables are entered into the regression process to predict achievement prior to the entry of expenditures, it is likely that both size and socioeconomic variables carry much of the correlation of expenditures with achievement into the calculation. The result is that the remaining correlation between expenditures and achievement is very small. In this case socioeconomic

factors carry most of the correlation into the calculation. The correlation between socioeconomic factors and expenditures is very high, pointing out the shared variance problem. It is a moot question as to why districts that serve students or families with low incomes have less money to spend on instruction and receive less state money for schooling.

Examining the way in which Walberg applies the regression models in several studies helps to explicate this discussion. Walberg appears to base the case contending no relationship between achievement and productivity on his theory of causal inferences on student learning and nine productivity factors (e.g., proficiency exams); on the triad relationship of socioeconomic status, productivity, and expenditures; and on his reliance on Hanushek's model and on early literature related to production function analysis.

In "A Guide to Educational Trouble-Shooting," Tyler (1984) praised Walberg's (1983) review of productivity research and his development of the theory, or paradigm perhaps, of school learning. Agreement is reached with Tyler in that the model appears to synthesize a large body of research in a clear and useful manner. Walberg and Fowler's (1989) model connects aptitude (ability, development, and motivation), instruction (amount and quality), and environment (home, classroom, peers, and television) as inputs to learning (affective, behavioral, and cognitive). It may be concluded that this model represents one accurate picture of a subset of variables that are precursors of productivity. Entry of these variables into a regression analysis results in the regression model being overspecified.

In Walberg's studies of district size, he permits urbanism to confound his variable. Walberg is frequently unclear regarding what is being measured as a variable representing productivity. The productivity variable often used is the percent of students passing a state-mandated proficiency examination. The method of measurement clearly restricts the range of the achievement construct and reduces the observed correlation. At other times, Walberg uses what he refers to as an efficiency measure, which consists of the predicted achievement score using psycho-economic status in the prediction equation divided by the observed achievement score. The configuration called "efficiency" appears to represent a measure of prediction error for psycho-economic status. Apparently his expenditure data include funds for transportation, lunch, special education, and similar programs that are not directly related to instruction.

Walberg fails to discuss how he handled the shared variance problem inherent in the triad relationship. His regression model enters psycho-economic status as the first predictor of students' test performances, size as the

second predictor, and, finally, expenditures as the third predictor. Little variance is left for expenditures to explain, after psycho-economic status and school size take credit for the shared variance. Psycho-economic status and expenditures have a large correlation. This correlation raises questions about the distribution of state funds.

Variable Specification

Variable specification is a problem. Confounding school factors such as special education, remediation processes, transportation costs, and lunchroom expenditures do not appear to be removed from the studies. It appears that the variable "expenditures" rather than "expenditures per student" are included in the correlation. The "efficiencies" prediction is still used as a dependent variable, and the truncated measure of productivity, percent passing, is used as a measure of achievement.

Multicollinearity

Often the search for controls leads to the overspecification of the model. When this happens, multicollinearity is most likely to occur. Such is the case reported by Fairley (1995) in the recent Pennsylvania equity hearings. In an attempt to use previous achievement measures as a proxy for intelligence, Fairley overspecifies the model that is being used to predict achievement from school expenditures. He reports that the multiple regression model shows a negative relationship between achievement and school expenditures, although the first-order correlation between these variables is positive. The sign reversal is a second evidence of multicollinearity, the first evidence being a singular matrix not permitting the completion of the analysis. The presence of multicollinearity promotes the misinterpretation of the results of the analysis.

A Global Across Time Comparison Model

Fairley (1995) moves from the simple correlation model of the production function analysis to the regression-based model of the production function analysis to the global across time model. His study was conducted for the Office of the Attorney General in Pennsylvania. In the simple linear production function for 501 districts, Fairley obtained a small, positive and statistically significant ($p < .05$) correlation suggesting a relationship between achievement and per pupil expenditures. Fairley then argued that when other correlates of schooling are controlled, the small correlation is wiped out. He added to his analysis personal income per district and fifth-grade achievement to predict eighth-grade achievement. Eighth-grade achievement is

then explained by way of regression by the following order of entry: average personal income per district, fifth-grade achievement, and per pupil expenditure. The shared variance problem occurs with the entry of personal income, and then this error in methodology is compounded by removing the relationship with eighth grade by the entry of fifth grade. Through this process any observable relationship between eighth-grade average achievement and average per pupil expenditures is successfully hidden in the residuals.

The global across time comparison utilizes a measure of educational expenditure across two specific years in history regardless of the elapsed time between the measures. Subsequently identified measures of educational achievement common to each of the two years described the changes across time. Fairley uses revenues per pupil nationally for 1972 and 1992. He finds that SAT mathematics and verbal score averages nationally exist as common measures of school achievement. Revenues per student increased from \$3,200 in 1972 to \$4,800 in 1992. During those 20 years, Fairley points out that revenues per student have increased by \$1,600. In 1972, SAT mathematics scale scores averaged 483, then decreased to an average of 476 in 1992. In 1972, SAT verbal average scale scores were 452 and decreased to 423 in 1992. Fairley concludes that clearly over this 20-year period, revenues are negatively associated with SAT performance and school achievement.

This type of reasoning leaves much room for criticism. Over the 20-year period, inflation has devalued the dollar approximately 30 percent; student participation in the SAT has increased, suggesting that more average students may be taking the examination; educational services may cost more since teachers salaries have increased approximately 18 percent; and the SAT has been updated and renormed one or two times over the period. It is hard for us to believe that this model explains anything except the difficulty of the question. This model has also been used by Hanushek (1996) to attack the meta-analysis by Hedges et al. (1994). Moreover, changes over time in the meaning of the variables were not questioned or even discussed.

The Threshold Driven Comparison Model

The last model discussed was developed in support of legal equity claims by school districts within specific states (Fortune & O'Neil, 1994). This model was developed to alter the analytic method of studying the problem in three ways. The changes sought by this model are: (a) to compare high and low expenditure groups as opposed to an association; (b) to overcome aggregation effects that are hiding differences in groups with regard

to expenditures; and (c) to permit the corrective use of concomitant variables. To address the comparison of the effects of large versus moderate expenditures, the research task appears to demand comparison rather than correlation. Rather than seeking to find the relationship across all districts in a state, it appears to be a better question to ask if differences exist within specific types of districts, such as those serving low-income versus high-income populations within their attendance areas or small versus large districts. The third change in analytic method is to accommodate the effects of aggregation and small differences in funding by establishing a purchasing threshold between the groups being compared. The model described and critiqued below has been used in support of court hearings in Alabama, Missouri, Ohio, Montana, South Dakota, and Rhode Island. The model has been used to provide information for groups of school districts in Virginia, New Hampshire, Pennsylvania, South Carolina, and Wisconsin. Money was demonstrated to make a difference in all states where the model was applied.

In many states there are several factors that threaten the consideration of the school districts as homogeneous elements or that mitigate against the expectation that spending patterns are similar. Size of districts, rural/urbanism, psycho-economics, and number of exceptional children (either gifted or at-risk) are variables whose subdivisions are likely to establish subsets of homogeneous groups with regard to expenditure requirements. In states such as Montana and Missouri, size is a dimension that creates homogeneous subgroups. In Alabama, rural/urbanism is the primary dimension that identifies homogeneous subgroups. In Ohio, income levels or psycho-economic status creates homogeneous subgroups. In some cases such as with Pennsylvania, there are one or two large, low-income urban districts which must be considered as outliers so as to establish homogeneous subgroups. In some states such as Rhode Island and South Dakota, analysis can be performed on the total number of groups in the states, and educational problems such as size of district or poverty do not appear to vary the cost of the instruction process across districts.

Homogeneous Group Comparisons

Creating pseudo-homogeneous subgroups of districts is another way of controlling for intervening variables that are also related to per pupil expenditures and/or student achievement. The intervening variables of income levels and school size can be used to create subgroups of districts so as to control for their effects, thus providing the control sought by the users of regression analysis through entering these variables into the analysis prior to the expenditure data. In using the

subgroup method of control for extraneous variables, one must be careful that group sizes do not become too small and that enough districts are included in the analysis so as to permit commentary on state practices. This proration can be subject to criticism of not using all the available variance in the data.

In 1970, a study conducted at Educational Testing Service (ETS) for the Office of Planning and Program Evaluation/Bureau of Elementary and Secondary Education/U.S. Office of Education (OPPE/BESE/USOE) (Herford, 1970) found that approximately \$300 per pupil was needed to improve elementary school children's reading scores by a one-month-of-training-experience level over the course of a year. A proration of this finding suggests a disparity of \$600 to \$700 is needed between districts. This was found by compounding the \$300 for 22 years at 3.5 percent. To create the analytic disparity within each homogeneous subgroup, the districts were ordered by instructional expenditures and were then divided into two groups where one is formed by the upper 30 percent and the other is formed by the lower 30 percent. The two groups are equal with regard to number of districts, and the funding level differences for most cases should exceed \$700. Given the satisfaction of these conditions, differences in achievement scores are likely to be observable.

Given the creation of the two groups (upper and lower 30 percent according to per pupil expenditures) from a single homogeneous subgroup and verification of a \$700 disparity, the independent *t*-test with pooled variance can be used to establish achievement test differences. If more than three homogeneous subgroups are to be analyzed, methods to deal with the inflation of the confidence level with multiple *t*-tests should be considered. Such methods include the recalculation of the confidence levels compensating for the use of several *t*-tests, recalculation of required probability levels using the Bonferroni techniques, or the use of the concept of families of *t*-tests, calculating an expected number of significant relations expected to be observed after considering the number of *t*-tests performed.

There is a question concerning the fact that all districts are not used directly in the comparison and concerning the appropriate subdivision of the groups. Should the districts be subdivided into 40 percent upper and lower subgroups? Or 20 percent upper and lower subgroups? These questions dismiss relationships which are being sought.

The threshold model has been used with state proficiency tests, norm-referenced national achievement tests, and statewide skill development tests. In each use of the model, positive relationships were found in the

DOES MONEY MAKE A DIFFERENCE?

state in which the model was applied. In Missouri, both district size and poverty appeared to serve as intervening variables. In South Carolina, the percentage of the people living in the district whose income fell below the poverty level served as an intervening variable. In South Dakota, size was the intervening variable. In the other states subgrouping by intervening variable was not necessary.

There are several bases for research criticism of this model. Many of these can be extrapolated from the criticisms of the correlational-based models. The achievement data must be in a metric that permits averaging, such as NCEs in achievement testing. The measurement of achievement in terms of percent passing or in terms of percentiles cannot be averaged and are therefore inappropriate for use in the model. Sometimes the above is the form in which states record achievement data, especially with regard to minimum competency tests. Data conversion must be made, and how much is lost by this conversion is unknown.

Homogeneity of districts with regard to educational spending is a concept similar in nature to a pinch of salt or a dash of pepper. Chefs can address this measure, but analytic description is difficult. Certainly, Philadelphia as a district is different from all others in Pennsylvania, as is St. Louis in Missouri and New York City in New York.

Because of their size and the complication of urban educational problems, these districts represent outliers with regard to educational spending. Usually, the model is applied to all districts in a state, and if the subgroups have a large variation or between group disparity in size, in levels of poverty, or in degrees of rural/urbanism, outliers are sought and subgroups are created on the active intervening variables.

Also, there is the question of the units of analysis for the model. Are the units of analysis "districts," or should there be weighting to account for student enrollment? What factors should be used to place limits on size of comparison groups? States that have less than 30 districts may already have groups too small for comparison. Should the groups be equal, or will "near-equal" suffice? Perhaps, one of the major concerns coming from Hanushek's (1989) work is how per pupil expenditure is used across districts within and across states. The translation of differences in the use of expenditures for resources seems to be a more interesting question for us to ask than those questions published about teachers making a difference.

The Structural Equation Model

This model is most likely a view of the future. The structural equation model has been in the research literature for two or more decades, but its use has been

limited, perhaps, because of the complexity of the model and the difficulty in interpreting the results given the present status of databases for certain problems. The benefits of the structural equation model include the ability to account for error and the ability to move beyond simple unidirectional paths of "causation." Wenglinsky (1996) used this model with the Common Core Data Base and the Teacher's Cost Index of the National Center for Education Statistics and the National Assessment of Educational Progress to address the question of how educational expenditures improve student performance and how they do not.

Although this application may be criticized with regard to the characteristics of the database and with regard to the cross-sectional nature of the study, it produces the most specific answers of the individual study attempts. The study by Wenglinsky addressed only the fourth and eighth grades and used a national database. Expenditures were studied by type, and Wenglinsky reported that money matters when it is spent on instruction. Differences in fourth- and eighth-grade results suggested that effects of money flow differently by grade level and suggested a need to study a larger range of grades.

What Is Known or Not Known

Two methods of research synthesis (e.g., summary of production function studies completed by the counting method and the meta-analysis of finance and achievement studies) have been used to show what is known. Both appear to point out different truths mainly due to what is not known and why the argument is being made. For instance, one does not know the degree to which the students included in many of the studies studied under the assumed funding conditions. They may have spent more than 75% of their educational experience in a school like the ones to which they are being compared. Furthermore, it is not known if the variance in funding in the research method represents the true variance in expenditures that are being studied. It is anticipated that the variance of expenditures within districts may exceed the variance between districts, since the variance between districts is based on aggregates. In addition to these unknowns, one does not know if the methods produce similar results for the same data because of the control problems and the levels problems. Logically, an investigator could argue that the unit of analysis should be students as opposed to districts. Arguments can also be made for units of analysis based on schools, grades, or even classrooms. Can a path analysis or a structural equation approach add to the sophistication of the research used to answer this important policy question? It appears from the single structural equation application that it will. However, it is important

to recognize that with this method too possible issues of variable selection and model specification may apply.

The Counting Method of Literature Summary

Although several other researchers and policy advocates have argued the question using the counting method of research summary, none have argued it as often, with as much vigor and persistence, in the light of as much controversy, and with as much influence on policy as has Hanushek (1971, 1981, 1986, 1989, 1992, 1994, 1996). It is interesting to read his prolific work, examine his contradictions in defense of his methodology as they evolve, and wonder at his persistence to his theme. As a critique to the counting method of research summary, a case study of Hanushek's work is warranted.

As Spencer and Wiley (1981) suggested, "Hanushek offers a provocative interpretation of the last two decades of research on educational productivity" (p. 46). Unfortunately, "Hanushek misinterprets the data on which he bases his conclusion and draws inappropriate policy implications from them" (p. 47). After reading a sampling of Hanushek's articles, one may concur with Hughes (1995) that one could quote from 20 years of Hanushek and destroy his current argument from previous years' quotes. However, it is preferable to examine his current thesis.

Hanushek (1989) contends that "Variations in school expenditures are not systematically related to variations in student performance," and that "... schools are operated in an economically inefficient manner" (p. 49). He suggests that "increased school expenditures by themselves offer no overall promise for improving education," and that "school decision making must move away from the traditional 'input directed' policies to ones providing performance incentives" (p. 49). Hanushek based this argument on the often-criticized study by Coleman et al. (1966), Equality of Educational Opportunity, and on his summary of 187 studies of education production functions. Hanushek's summary is a continuing theme throughout his publications. The summary process began in 1981 with 29 articles and 130 studies. It was continued in 1986 with 33 articles and 147 studies, and it culminated in 1989 with 38 articles and 187 studies. The summary serves as the research foundation for supporting Hanushek's conjecture of no relationship between school districts' expenditures and student performance on standardized achievement tests.

There are several serious omissions and research flaws in the description and logic of the Hanushek summaries. These include the lack of disclosure of sample sizes in the studies that were reviewed, inadequacy in size and representativeness of the 187 case studies, misinterpretation of the results of the hypothesis testing, potential

misinterpretation of the summary, failure to use selected research that is not consistent with the ideas being promoted, and inadequate specification of key study variables.

It appears that the primary criterion for qualification of the studies reviewed by Hanushek was publication. Hanushek states that at least one study deals with a district or districts in all regions of the United States, with different grade levels, and across different performance measures. He provided two tables to describe the sample. In Table 1 of his 1989 article, "The Impact of Differential Expenditures on School Performance," Hanushek showed the number of studies dealing with single districts (60) and the number dealing with multiple districts (127), but he failed to provide any information on the number of districts involved in the multiple districts.

While Hanushek has failed in regard to sampling requirements, the size problem may be less bothersome than the lack of randomness. The 187 studies were probably conducted in reaction to some problem or needed inquiry. The nature of the studies suggest that the problem is likely involved with school finance. These unusual districts are potentially different from the remaining districts in the population. No evidence is presented to suggest that generalizability is possible.

Former Secretary of Education Bennett referred to production function studies on several occasions. On February 26, 1988, Bennett remarked, "Money doesn't cure school problems." On February 29, 1988, he was more explicit, "We've done 147 studies at the Department of Education. We cannot show a strong, positive correlation between spending more and getting a better result." In an earlier reference to those studies, he had said on April 13, 1987, that "in only two or three do we find even a weak correlation between spending and achievement." The 147 studies referred to by Bennett are those summarized by Hanushek using the production function process.

Baker (1991) discussed Hanushek's absence of a decision rule in his summary of the literature for the 147 studies. Baker stated that synthesis of literature as reported by Hanushek can be conducted in one of two ways: either by the vote-counting method with stated expectancy (decision rule) or by the meta-analysis method. Hanushek does not compute effect sizes, leaving his review to the vote-counting method. Given the absence of the statement of a decision-rule by Hanushek, Baker assumes a decision-rule that 5% of the studies will be significant by chance. He then showed that 20% of the studies were significant; 75% of the 65 studies relating expenditures to student performance were found not significant; 87% of the 61 studies relating administrative inputs to student performance were not found significant;

and 84% of the 74 studies relating facilities to student performance were not found significant.

Meta-Analysis in the Address of the Question

Four applications of meta-analysis related to the address of the question have appeared in the literature over the last decade. One analysis by Child and Shakeshaft (1986) covered only 45 studies and showed that expenditures spent directly on instruction made a difference. Two of the analyses used the same sample of studies summarized by Hanushek (Greenwald, Hedges, & Laine, 1994; Hedges, Laine, & Greenwald, 1994). The studies included in the meta-analyses reported by these three researchers appear in the appendix of the article "The Effect of School Resources on Student Achievement," in the *Review of Educational Research* (1996). The Hedges et al. (1994) article was briefer and addressed only the question of a relationship between student achievement and per pupil expenditures. The Greenwald et al. (1994) article addressed the question and pointed out fallacies in Hanushek's summary. In the latter two articles the researchers found significant ($p < .05$) positive effect sizes in contrast to the picture of inconclusion painted by Hanushek. The differences in the results is likely due to the capability of the summary process to take into account the magnitude of effects reported in the results of the summarized studies, the requirement by the method to be more stringent in the qualifying of studies to be included in the analysis, and the requirement by the method to address more systematically variable specification.

Laine, Greenwald, and Hedges (1995) provide another meta-analysis based on a review of production function results using an expanded universe of studies. Effect sizes were analyzed for seven inputs, and results were summarized for the total sample, for studies completed before 1970 and for studies conducted after 1970. The break using 1970 is important for two reasons. First, the studies before 1970 were likely to have been conducted for a reason other than the study of the relationship between money and achievement. Second, 1970 can be argued to close one economic era and begin another. Regardless, the researchers concluded "that school resources are systematically related to student achievement and that these relations are large enough to be educationally important" (p. 57).

The meta-analysis process is certainly more systematic and less arbitrary than the counting method, but again the process is only as "good" as the studies available for summary. Many aspects of the success of this method depends on the thoroughness of the researchers using it. How rigorous are they in finding

"all" of the relevant studies? How systematic are they in qualifying the studies? How appropriate are their views of policy relevance?

In addition to the model's dependence on the rigor of the researchers, the model is also dependent on the characteristics of what is out there. Are the summarized studies dispersed so as to provide representation to the population and not overrepresentation to a single grade level, region, and type of district? One suspects that more studies are on fourth grade, eighth grade, and tenth grade because of school districts testing at those grade levels. Variable specification and the effects of the reason for the study must be controlled for by the researchers using the meta-analysis.

A Policy Position for Today

Based upon evidence provided in the relevant literature, a nontrivial relationship between per pupil expenditures and student achievement test performance seems apparent (Fortune & O'Neil, 1994, Lockwood & McLean, 1993, 1997). If the evidence is truly inconclusive, then we believe that we should face the uncertainty in a similar manner, as did Blaise Pascal. Pascal in his *Pensées* speaks of believing in God. "The indifference of the skeptic," Pascal wrote, "is to be overcome by means of the 'wager.'" If God does not exist, then the skeptic loses nothing by not believing in him, but if he does exist, the skeptic gains eternal life by believing in him. If educational expenditures are not systematically related in a policy important manner to student achievement, then according to our evidence only modest gains will be made in some cases. However, if educational expenditures are systematically related in a policy important manner to student achievement, we have a powerful tool to improve education. Lockwood and McLean (1993, 1997), suggest that there may be a curvilinear relationship between funding and achievement (i.e., threshold 1—buildings, utilities, chalk, etc. and threshold 2—monies used to directly improve the educational process) that should be taken into consideration and investigated as well.

A Policy Position for the Future

To reach the important research goal of answering the question on policy making, the following need to be accomplished: investigating the assumptions underlying past attempts to answer the policy question; investigating the levels problem; developing specifications for data definition, storage, and processing so as to eliminate the measurement problem; and comparing the five methods using the same data. The assumptions include duration of educational experience of students included in the study,

estimation of the within district variance in expenditures, definition of controls needed, and satisfaction of the assumptions of the statistical model in use. The levels problem involves the estimation of the effects of data aggregation and the development of guidelines for the appropriate choice of units of analysis. Measurement issues involve attention to time of data collection, accuracy of reporting, guidelines for the construction of variables (such as use of Average Daily Membership (ADM), Average Daily Attendance (ADA), weighted ADM, or enrollment in per pupil expenditure calculation), building of longitudinal files, tracking of students, and the tracking of costs. The comparison of the methods for analysis using the data from the same school districts (using schools as the unit of analysis) to compare the linear correlation method, the multiple regression method, the over time direct comparison method, the threshold method, and the structural equation method should explain the reported differences in results that cannot be explained by limitations of the data or convictions of the researchers. Structural equation applications to a more comprehensive database appear to be the next step.

References

- Baker, K. (1991). Yes, throw money at schools. *Phi Delta Kappan*, 72(8), 628-631.
- Beiker, R. F., & Anschel, K. R. (1973). Estimating educational production functions for rural high schools: Some findings. *American Journal of Agricultural Economics*, 55, 515-519.
- Boardman, A., Davis, O. A., & Sanday, P. R. (1977). A simultaneous equations model of the education process. *Journal of Public Economics*, 7(1), 23-49.
- Bridge, G. R., Judd, C. M., & Moock, P. R. (1979). *The determinants of educational outcomes: The impact of families, peers, teachers and schools*. Cambridge, MA: Ballinger.
- Brown, B. W., & Saks, D. H. (1975). The production and distribution of cognitive skills within schools. *Journal of Political Economy*, 83(3), 571-593.
- Burkhead, J. (1967). *Input-output in large city high schools*. Syracuse, NY: Syracuse University Press.
- Child, T. S., & Shakeshaft, C. (1986). A meta-analysis of research on the relationship between educational expenditures and student achievement. *Journal of Educational Finance*, 12, 249-263.
- Cohn, E. (1975). Economies of scale in Iowa high school operations. *Journal of Human Resources*, 3(4), 422-434.
- Coleman, J. S., Campbell, E. Q., Hobson, C. F., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Department of Health, Education and Welfare.
- Fairley, W. B. (1995). *Spending and economic base in Pennsylvania school districts*. An unpublished paper for the Pennsylvania Office of Attorney General, Swathmore, PA.
- Ferguson, R. H. (1992). Paying for public education: New evidence on how and why money matters. *Harvard Journal of Legislation*, 28, 483-397.
- Fortune, J. C., & O'Neil, J. S. (1994). Production function analyses and the study of educational funding equity: A methodological critique. *Journal of Educational Finance*, 20(1), 21-46.
- Greenwald, R., Hedges, L. V., & Laine, R. D. (1994). When reinventing the wheel is not necessary: A case study in the use of meta-analysis in educational finance. *Journal of Educational Finance*, 20, 1-20.
- Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66(3), 361-396.
- Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro-data. *American Economic Review*, 61, 280-288.
- Hanushek, E. A. (1981). Throwing money at schools. *Journal of Policy Analysis and Management*, 1(1), 19-41.
- Hanushek, E. A. (1986). The economies of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24, 1141-1177.
- Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher*, 16(4), 45-51, 62.
- Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy*, 100, 84-117.
- Hanushek, E. A. (1994). Money might matter somewhere: A response to Hedges, Laine, & Greenwald. *Educational Researcher*, 23, 5-8.
- Hanushek, E. A. (1996). School resources and student performance. In G. Burtless (Ed.), *Does money matter?* (pp. 43-73). Washington, DC: The Brookings Institution Press.
- Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money Matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23(4), 5-14.
- Herford, K. (1970). Informal memo evaluation of the ESEA Title I OPPE/BESE.
- Hughes, M. (1995). Liddel v. State Department of Education St. Lois, MO. Paper commissioned by

DOES MONEY MAKE A DIFFERENCE?

- Murray State, Murray State, KY. Case No. 72-0100C(6).
- Laine, R. D., Greenwald, R., & Hedges, L. V. (1994). Money does matter: A research synthesis of a new universe of education production function studies. In L. O. Pincus & J. L. Wattenbarger (Eds.), *Where does the money go? Resource allocation in elementary and secondary schools*. Thousand Oaks, CA: Corwin.
- Lockwood, R. E., & McLean, J. E. (1993, November). *Educational funding and student achievement: You be the judge*. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, LA.
- Lockwood, R. E., & McLean, J. E. (1997). Twenty-five years of data on educational funding and student achievement: What does it mean? *Educational Research Quarterly*, 21(2), 3-11.
- Raymond, R. (1968). Determinants of the quality of primary and secondary public education in West Virginia. *Journal of Human Resources*, 3(4), 450-470.
- Spencer, B. D., & Wiley, D. E. (1981). The sense and nonsense of school effectiveness. *Journal of Policy Analysis and Management*, 1(1), 43-52.
- Tuckman, H. P. (1971). High school inputs and their contributions to school performance. *Journal of Human Resources*, 6(4), 490-509.
- Tyler, R. (1984). A guide to educational troubleshooting. *Educational Leadership*, 41(8), 27-30.
- Walberg, H. J. (1983). Variables in measuring educational productivity. *Symposium of Educational Productivity: The Proceedings, October 23, 1983*. Austin, TX.
- Walberg, H. J., & Fowler, Jr., W. J. (1989). *Expenditures and size efficiencies of public school districts*. (ERIC ED 274 471, RC 015786)
- Wenglinsky, H. (1996). *When money matters*. Princeton, NJ: Educational Testing Services, Policy Information Center.
- Winkler, D. R. (1975). Educational achievement and school peer group composition. *Journal of Human Resources*, 10(2), 190-203.

An Analysis of Factors Related to School-District Size in Arkansas

Michael B. Gilbert
Central Michigan University

This project examined and analyzed certain critical factors concerning school-district size in Arkansas. These factors include: standardized test scores, per-pupil expenditures, remediation rates, and percentages of students eligible for free and reduced-price meals. Multiple correlations and regression analyses were performed. Two questions guided the study: (1) Is there a minimum school-district size that can be identified as functional in Arkansas? and (2) How is school-district size related to input factors of per-pupil expenditures and output factors of standardized test scores and remediation rates? The data did not reveal a functional relationship between school-district size and the other variables. There was statistical significance demonstrated at low levels, attributable more to the size of the data base than to anything else.

Introduction

The issue of learning environments has been discussed and "cussed" for many years in and out of educational circles. What is a "good" class size? How big (or small) should a school be? What is the optimal size of a school district?

Although some have proposed answers to these questions, there is not uniform agreement. In fact, many unrational (possibly, irrational) factors may intrude into the argument(s) (e.g., community tradition, political expediency, etc.). This study was undertaken to attempt to derive a reasonable answer to the question of school-district size as related to specific output factors.

Description of the Project

In Arkansas the 311 public school districts averaged about 1,400 pupils in average daily attendance in 1994-95, ranging from fewer than 100 to more than 25,000. Eighty-seven percent of the districts had fewer than 2,500 pupils, less than one half of the minimum district size suggested in much of the facilities planning literature; 64% of the districts had fewer than 1,000 pupils.

Other neighboring states, with the exception of Mississippi, averaged more than the 5,000 pupils, considered to be a reasonable minimum district size, considering pupil-teacher ratios, district organization, and feeder patterns. The district averages for other states in the region are listed in Table 1. For the nine states, there was an average of 143 districts per state, with a mean of almost 6,000 students per district.

Michael B. Gilbert is Professor of Educational Leadership at Central Michigan University. Correspondence regarding this article should be sent to Michael B. Gilbert, Central Michigan University, 314 Ronan Hall, Mt. Pleasant, Michigan 48859 or via e-mail at Michael.Gilbert@cmich.edu.

Table 1
Average School District Size in Selected
Southern States, 1994-95

State	No. of Districts	Avg. Size
AR	311	1,400
GA	187	6,000
MS	154	3,200
TN	141	5,800
NC	140	7,800
AL	130	5,600
SC	92	6,600
FL	67	24,000
LA	66	12,000
Average	143	5,975

All of the states have "unified" school districts; that is, each district encompasses K-12 grades. Some states use existing political subdivisions (counties/parishes) to define the districts (with some exceptions) (e.g., Georgia, Florida, and Louisiana). Most of the other states have no discernible pattern. Arkansas has by far the largest number of school districts and the smallest average number of pupils. Arkansas has two thirds more districts than the next closest state in number of districts (Georgia) and less than one half of the next closest average (Mississippi). More than one half of the districts in Arkansas have fewer than 750 pupils. [Previous data have supported the conclusion that the "ideal" district in Arkansas has between 800 and 9,000 students (Calvery et al., 1994).]

While these data seem to compel examination and the situation seems to need to be addressed, the data are not clear as to whether small districts and small schools are better or worse. The Arkansas Department of Education specifies two indicators of *distress* for schools/districts -

academic and financial. Of those entities designated to be in distress, the patterns are not clearly related to school or district size.

Arkansas, while relatively small geographically, is sparsely populated. [Arkansas is the same size as Great Britain, with only four percent of the British population.] Pupil population is dispersed among the relatively large number of school districts. This causes both administrative and instructional considerations and concerns. If a reasonable pupil-teacher ratio is 20-25:1 and pupil assignment should have some options in matching pupils with teachers, then at a minimum a six-grade elementary school should have 12 teachers. These teachers would serve a pupil population of 240-300. [Students in medium-sized (more than 300) schools have been shown to achieve higher scores in reading and mathematics than in smaller schools (University of Michigan, 1996).] If the secondary grades need the same numbers of teachers at the same ratio, then about one third of the districts in the State would fall short of the extrapolated minimums.

These capacities echo reference to the criteria suggested by the National Council on Schoolhouse Construction: elementary schools "should range between 300 and 500 students. At the secondary level, the minimum size is set at 300 students and the maximum at 1,200. These enrollment ranges reflect the views of many educators" (Castaldi, 1994, p. 139). These figures harken back to Englehardt (1970, p. 3), who said that "it can be determined fairly objectively that a school of 350 to 400 pupils is a minimum size." Some criteria for planning schools suggest absolute minimum sizes: elementary (K-5), 300; middle (6-8), 450; and senior high (9-12), 1000 (Atlanta Public Schools, 1972, p. V.1). "Dysfunctional" schools in South Carolina are substantially smaller when compared to those designated as "award-winning" (Stevenson, 1996).

If there were not enough options for instruction, then outcomes might suffer. That is, pupils would be hard-pressed to meet expectations. Pupils least like their teachers perform poorest (Gilbert, 1994). This reinforces the idea that grades are a reflection of teacher expectations. Opportunities for pupils to be matched with teachers would increase if there were more teachers available.

The issue of more teachers relates to class size and may have substantial financial consequences. A longitudinal study in Tennessee showed that reducing class size had discernible and important effects on learning (Achilles, Finn, & Bain, 1997-98; Bracey, 1995; Viadero, 1995). However, class size alone did not influence learning. Other strategies and considerations needed to be in place as well (e.g., improved teaching styles and improved teaching materials) (Hoff, 1998; Prais, 1996).

Purpose

This project examined and analyzed certain critical factors concerning school-district size to determine if there were specific points at which districts become functional or dysfunctional. These factors included: standardized test scores, remediation rates for students entering college, and per-pupil expenditures.

Significance of the Project

Arkansas consistently has among the lowest educational outcomes of the United States. While Arkansas test scores have risen slightly over the past several years, other data have continued near the bottom (e.g., per-pupil expenditures).

Some key comparisons show that Arkansas teachers average less than 80% of the salaries of their national counterparts. School districts in Arkansas spend about 70 percent of the national per-pupil average (Education Vital Signs, 1997). An additional consideration is that the quality of the educational programs in Arkansas may be measured, in part, by those who attend college – approximately one half of those who attend public colleges and universities need remediation in their first year. This means that the programs from which these students have emerged have not prepared them adequately for higher-education academics. Further, the retention rates for those needing remediation is below 60% at the state's metropolitan university, the University of Arkansas at Little Rock (UALR, 1997). Not only are students arriving poorly prepared for academic work, but the remediation may not be sufficient to upgrade their skills and knowledge for continuation in a degree program.

A critical political question has been school-district consolidation. If there are too many school districts, what is the most logical (palatable?) alternative configuration? Every time reducing the number of school districts has been broached in a political forum, the smaller districts muster an army of detractors without much more than invoking tradition and history: "This is the way we've always done things. Why should we change?" Those who have proposed the issue have not had the ammunition to counter, other than an opposing opinion (Gilbert, 1989a, 1989b). No evidence has been presented that consolidating small school districts into larger ones will necessarily change outcomes (Adams, 1994; Howley, 1997), namely, standardized test scores and dropout rates (Kennedy, 1989). These conclusions continue to echo previous studies: Consolidation has not alleviated the educational problems endemic to rural areas (Sher, 1977), and Arkansas is largely a rural state. Neither financial input nor district size has shown any significant positive student outcomes (Alkin, 1968).

Student achievement has been shown to be inversely related to district size (Walberg & Walberg, 1994). In fact, smaller schools and districts may be more efficient in enhancing educational outcomes than larger ones (Fowler & Walberg, 1991). The variable of socioeconomic status (SES) may enter the "equation" – smaller districts have been shown to yield higher achievement when SES and per-student expenditures are taken into account (Walberg, 1989; Walberg & Fowler, 1986).

The middle and later parts of this century have seen the number and composition of school districts in this country change dramatically. The number of districts has decreased some 87% (from more than 117,000 in 1940 to fewer than 16,000 currently, approximately one fourth of which are in California, Illinois, Nebraska, and Texas), while the average district enrollment has increased more than tenfold (from fewer than 200 in 1940 to more than 2,600 currently). The number of public schools has shrunk by almost 70% (from approximately 200,000 to fewer than 63,000). Although many of these changes were occasioned by concern for financial efficiency, the results of these (political) decisions have not fostered salutary outcomes – average state scores are significantly and negatively related to average school size, average district size, and percentage of district funding paid by the state (Walberg, 1994).

There is a dichotomy of purpose which continues to fuel debate on district and school size – larger districts can provide a wider variety of curricular choices, materials and facilities, and smaller districts can provide a more personalized learning environment (Witcher, 1997). Collateral to these issues is local control (Walberg & Walberg, 1994). What is the desirable locus for district decision making and how should community be defined? That is, do we continue to look at small geographic areas as the communities that define school districts, or is there a broader perspective? The sense of organizational and community culture may explain, in part, why small schools and districts do well (Swanson, 1991), especially in rural areas. School-district size may be an elusive variable because the nature and mission of school districts are not uniformly defined (Howley, 1996; Medrich, 1994). Further, with more outcome measures available and the technology for analyzing them at hand, states may not need to address (mandate) district size and other input variables (Monk, 1994).

"Hard" data will not compel change without other considerations, but empirical research might provide the key decision makers with a careful analysis of the situation and recommendations for change based on the research. This was the underlying notion related to the significance of the project – to provide a research-based

examination of critical factors related to school-district size as those factors might impinge on future decisions.

Related Information

The "Western Electric study showed conclusively that environment did make a difference" (Sommer, 1969, p. 165); hence, the *Hawthorne effect* was posited. Various anthropological aspects of space led to the notions of *proxemics* – how close people are comfortable with each other in social situations (Hall, 1969). What can be derived from both these observations is that space is an important component in interaction and productivity. There must be enough space to promote comfort and facilitate output. Too little space or other constraints on the environment might inhibit salutary program results. Total space in schools is a function of self-contained teaching stations (Arkansas Department of Education, 1992) – that is, the number of teachers and the subjects they teach. By extension, and related to outputs such as need for remediation in college, an insufficient number of staff and subjects might have deleterious – or at least, not fully productive – results.

Project Plan and Methodology

The design of the study included examination of the focal variables collected statewide in Arkansas. Data were available from the Arkansas Department of Education (1998) to provide the information with which the study was conducted. The research questions were:

- Is there a minimum school-district size in Arkansas that can be identified as functional?
- How is school-district size related to input factors of per-pupil expenditures and output factors of standardized test scores and remediation rates in Arkansas?

Data were analyzed using multiple correlations and multiple regression analyses. The focus was to determine if there is a point at which a school district becomes less productive. That is, is there a predictable size for school-district functionality?

Results

With regard to school-district ($N = 311$) SIZE, there were several significant correlations, but all were inconsequential ($r < .21$). They are shown in Table 2.

Because of the low correlations there was nothing compelling to report, but it was noted as *district size increased*, (a) the percentage of students qualifying for free/reduced-price meals decreased, (b) more money was generated per student, (c) there were higher percentages of students at or above the 75th percentile on standardized tests, (d) higher percentages of students took the ACT,

and (e) higher percentages of students scored at 19 or above on the ACT.

Table 2
Significant Correlations of Selected Variables
with School-District Size

Eligibility for free/reduced-price meals	-.19**
Local financial effort	.18**
Scores at or above 75th percentile	.20**
Percentage of seniors attempting the ACT	.16**
Percentage of ACT scores at or above 19	.13*

* $p < .05$; ** $p < .01$

With regard to percentages of students needing REMEDIATION in their first year at public IHE's, the data became somewhat more interesting in the area of significant findings. These data are shown in Table 3.

Table 3
Significant Correlations of Selected Variables
with Remediation Rates

Dropout rate	.12*
Percentage of standardized scores at or below the 25th percentile	.28**
Percentage of students eligible for free/reduced-price meals lunch	.32**
Percentage of standardized scores at or above the 75th percentile	-.31**
Percentage of ACT scores at or above 19	-.36**

* $p < .05$; ** $p < .01$

As the percentage of students needing remediation their first year increased, (a) there was a slight positive correlation with the district dropout rate (however, that is a grades 7-12 calculation), (b) higher percentages of total students scored at the 25th percentile or lower, (c) higher percentages of total students qualified for free/reduced-price meals, (d) lower percentages of students scored at the 75th percentile or higher (or lower percentages of students needed remediation as there are higher percentages of students at or above the 75th percentile, the likely pool for college attenders), and (e) similarly, lower percentages of students needed remediation as the district percentages of students scoring at or above 19 on the ACT rose.

In a regression analysis, the best indicator of students needing remediation was lower district percentages of students with ACT scores at or above 19, followed by students coming from districts with lower percentages of standardized test scores at or above the 75th percentile, then lower average ACT scores. So if admissions officers were looking for recruitment factors, they might focus on

students with higher ACT scores, who would be coming from Arkansas districts with higher percentages of standardized scores at or above the 75th percentile. Table 4 shows the hierarchical array of variables predicting the need for REMEDIATION.

Table 4
Summary of Hierarchical Regression Analysis for Variables
Predicting the Need for Remediation of First-Year Students
at Public Colleges and Universities in Arkansas

Variable	B	SE B	B
Step 1			
ACT 19 or higher	-.42	.06	-.36***
Step 2			
ACT 19 or higher	-.32	.07	-.28***
Std. scores at or above 75th %ile	-.56	.18	-.18**
Step 3			
ACT 19 or higher	-.33	.07	-.28***
Std. scores at or above 75th %ile	-.54	.18	-.17**
Avg. ACT scores	-.02	.01	-.11*

Note. The B scores represent the regression weights for each variable; the SE B scores represent the standard error for each B score; the Beta (B) scores are the regression weights by which all of the variables are represented in standard score form.

* $p < .05$; ** $p < .01$; *** $p < .001$.

With regard to PER-PUPIL EXPENDITURES, there are related indicators. Two seem to be compelling: eligibility for free/reduced-price meals and local financial effort. These data are shown in Table 5.

Table 5
Significant Correlations of Selected Variables
with Per-Pupil Expenditures

Percentage of standardized scores at or below the 25th percentile	.12*
Percentage of students eligible for free/reduced-price meals	.30**
Local financial effort	.48**
Percentage of standardized scores at or above the 75th percentile	-.17**
Percentage of seniors attempting the ACT	.22**
Percentage of students at or above 19 on the ACT	.17**

* $p < .05$; ** $p < .01$

Districts that spend more money per pupil were characterized by: (a) having higher percentages of students qualifying for free/reduced-price meals, (b) generating more money locally, (c) having higher percentages of students taking the ACT and scoring at 19 or above, and (d) having higher percentages of students scoring at or below the 25th percentile and lower percentages of students

scoring at or above the 75th percentile. What appears to be interesting here is that *more money did not appear to result in better overall district performance*. Districts with poorer students spend more, but more money is not producing salutary results, as defined in this study.

Standardized test score results were predicted predominantly by eligibility for free or reduced-price meals. Percentages of low scores (25th percentile or below) were also predicted by dropout rates and remediation rates. Percentages of high scores were also predicted by percentages of district ACT scores of 19 or higher and district financial effort. Tables 6 and 7 show the hierarchical arrays of factors.

Table 6

Summary of Hierarchical Regression Analysis for Variables Predicting the Percentages of Low Standardized Test Scores in Public School Districts in Arkansas

Variable	B	SE B	B
Step 1			
Free/reduced-price meals	.30	.03	.49***
Step 2			
Free/reduced-price meals	.28	.03	.46***
Dropout rate	.43	.15	.14**
Step 3			
Free/reduced-price meals	.26	.03	.42***
Dropout rate	.41	.15	.14**
Remediation rates	.06	.02	.13*

Note. The *B* scores represent the regression weights for each variable; the *SE B* scores represent the standard error for each *B* score; the Beta (*B*) scores are the regression weights by which all of the variables are represented in standard score form.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 7

Summary of Hierarchical Regression Analysis for Variables Predicting the Percentages of High Standardized Test Scores in Public School Districts in Arkansas

Variable	B	SE B	B
Step 1			
Free/reduced-price meals	-.23	.02	-.55***
Step 2			
Free/reduced-price meals	-.18	.02	-.42***
ACT 19 or higher	.09	.02	.25***
Step 3			
Free/reduced-price meals	-.17	.02	-.41***
ACT 19 or higher	.09	.02	.25**
Total local financial effort	.001	<.01	.10*

Note. The *B* scores represent the regression weights for each variable; the *SE B* scores represent the standard error for each *B* score; the Beta (*B*) scores are the regression weights by which all of the variables are represented in standard score form.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Discussion

The study was initiated because of questions associated with the unusually high number of small districts in Arkansas. The presumption was that smaller districts were impacted negatively in their ability to produce positive results. *That presumption was not borne out*. What is apparent is that districts with poorer students do not fare well, regardless of the amount of money they spend (cf. Lamin, 1995; Schmuck & Schmuck, 1992). This finding echoes previous work that consolidation in Arkansas would not necessarily produce better outcomes (Kennedy, 1989), and that school-district size has not been found to be a statistically significant factor of student achievement (Witcher, 1997).

To answer the questions that guided the study:

- *Is there a minimum school-district size that can be identified as functional in Arkansas?* The data did not reveal a functional relationship between school-district size and the other variables. There was statistical significance demonstrated at low levels, attributable more to the size of the data base than to anything else. The interesting relationships in need of further examination regarding school-district size were with the following variables: (a) eligibility for free/reduced-price meals (the larger the district the lower the percentage), (b) local financial effort (the larger the district the greater the effort), and (c) standardized scores at or above 75th percentile (the larger the district the higher the percentage).

- *How is school-district size in Arkansas related to input factors of per-pupil expenditures and output factors of standardized test scores and remediation rates?* There does not appear to be a direct relationship, other than the slight positive correlation of higher percentages of standardized test scores with district size.

What implications exist for school districts to change patterns of performance? Many years ago when asked by Dick Cavett why the prison system in Georgia was in such abysmal shape, former governor Lester Maddox replied that the situation would be better if there were a better class of prisoners. Districts might do better if they had "better" students to work with at the outset.

Conclusions and Recommendations

What we have is another example of schools trying to deal with a situation one step removed from reality. Desegregating schools is easier where there are desegregated neighborhoods, but we cannot legislate neighborhood composition. Similarly, it seems that students performing better academically come from more affluent homes (as one of my colleagues has said this is an example of a BFO [blinding flash of the obvious]). We cannot assure (or

control) that kids will be born to parents who are educated, who love them and want them to succeed, and who support their efforts in school by providing a suitable home environment.

Arkansas is a relatively poor state in terms of per-pupil expenditures, teacher salaries, per-capita income, and other related data. It is a state with high levels of teenage pregnancy and low levels of educated adults. All of the surrounding and regional states – with the possible exception of Mississippi (the next closest in average school-district size) – have higher levels of educational inputs and outcomes. If it is apparent that more money spent on poorer children does not provide them with an effective education – that is, one that will prepare them to compete successfully once they leave the public schools – then what other strategies should be explored? Or is the question what is being done with the extra money? Are categorical compensatory dollars spent in only certain, externally-defined ways? Or is money being spent on things that do not relate to actually compensating for the deficiencies that poorer children bring with them to school (e.g., better teachers, better teaching strategies, better learning materials, etc.)?

These are knotty questions drawn from a complex problem. Public-school districts do not start with “raw” materials. The units of focus – the students – have been impacted by many things prior to their entrance into school. The schools can only control so much; the rest is a product of other environmental, social, and family factors.

School-district size and school size have been variables in need of definitive empirical analysis in order that they compel (insofar as possible) rational decision making with regard to educational outputs and results. Educational planning may use established benchmarks to guide school-district efforts to create a suitable environment for learning, but there may be other mitigating factors to be considered – history, tradition, organizational culture, etc.

This and other studies have raised questions in need of further address. For example, if larger districts and larger schools (within suggested ranges) provide greater opportunities for learning, how does one explain the lack of conclusive evidence found in this study (other than by speculation)? What other factors may intervene in determining the effect of size on results? Or it may be that earlier consensus is still accurate – small schools are more effective, encourage community and school identity, and relate to individual fulfillment and participation. Conversely, large schools are more inefficient, have an air of institutional bureaucracy, and have greater degrees of personal loneliness (Ornstein, 1990).

These questions may provide the direction for future studies, since some questions have been answered but not conclusively. Also, designs similar to the one used in this

study might be considered for replication using data from other states and regionally.

References

- Achilles, C. M., Finn, J. D., & Bain, H. P. (1997, December-1998, January). Using class size to reduce the equity gap. *Educational Leadership*, 40-43.
- Adams, J. E., Jr. (1994). *School district size and state educational costs in Kentucky* (Revised). (ERIC Document Reproduction No. ED371921)
- Alkin, M. C. (1968, February). *Economy of scale in the production of selected educational outcomes*. Paper prepared for the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction No. ED022239)
- Arkansas Department of Education. (1992). *Rules and regulations: Minimum schoolhouse construction standards*. Little Rock, AR: Author.
- Arkansas Department of Education. (1998). *District report card*. Little Rock, AR: Author.
- Atlanta Public Schools. (1972). *Educational facilities survey*. Atlanta, GA: Author.
- Bracey, G. W. (1995). Research oozes into practice: The case of class size. *Phi Delta Kappan*, 77, 89-90.
- Calvery, R. et al. (1994). *A comparison study of the characteristics of the ten best school districts in the state of Arkansas for 1992-93*. (ERIC Document Reproduction No. ED382627)
- Castaldi, B. (1994). *Educational facilities: Planning, modernization, and management*. Boston: Allyn & Bacon.
- Education Vital Signs. (1997). *American School Boards Journal*, 184(12), A26-31.
- Englehardt, N. L. (1970). *Complete guide for planning new schools*. West Nyack, NY: Parker Publishing.
- Fowler, W. J., & Walberg, H. J. (1991). School size, characteristics, and outcomes. *Educational Evaluation and Policy Analysis*, 13(2), 189-202. (ERIC Document Reproduction No. EJ430692)
- Gilbert, M. B. (1989a, February 23). Countywide school districts make sense for the whole state. *Arkansas Gazette*, p. 7B.
- Gilbert, M. B. (1989b, November 13). School numbers add up to too many districts. *Arkansas Gazette*, p. 5B.
- Gilbert, M. B. (1994). *Meeting communication needs of students can promote success*. Unpublished off-campus duty assignment report, University of Arkansas at Little Rock.
- Hall, E. T. (1969). *The hidden dimension*. Garden City, NY: Anchor Books.
- Hoff, D. J. (1998, October 7). Scholars seek new solutions to the “achievement gap.” *Education Week*, p. 5.

SCHOOL-DISTRICT SIZE

- Howley, C. (1996). *Ongoing dilemmas of school size: A short story*. (ERIC Document Reproduction No. ED401089)
- Howley, C. (1997). Dumbing down by sizing up. *School Administrator*, 54(9), 24-26, 28, 30. (ERIC Document Reproduction No. EJ551985)
- Kennedy, R. L. (1989). *Site, expenditures, MAT6 scores, and dropout rates: A correlational study of Arkansas school districts*. (ERIC Document Reproduction No. ED361160)
- Lamdin, D. J. (1995). Testing for the effect of school size on student achievement within a school district. *Education Economics*, 3(1), 33-42. (ERIC Document Reproduction No. EJ505007)
- Medrich, E. A. (1994). *Size, cost, and quality of schools and school districts: A question of context*. (ERIC Document Reproduction No. ED361162)
- Monk, D. H. (1994). *Modern conceptions of educational quality and state policy regarding small schooling units*. (ERIC Document Reproduction No. ED303910)
- Ornstein, A. J. (1990). School size and effectiveness: Policy implications. *Urban Review*, 22(3), 239-45.
- Prais, S. J. (1996). Class size and learning: The Tennessee experiment— what follows? *Oxford Review of Education*, 22, 399-414.
- Schmuck, R. A., & Schmuck, P. A. (1992). *Small districts, big problems: Making school everybody's house*. (ERIC Document Reproduction No. ED370747)
- Sher, J. P. (1977). *Economy, efficiency, and equality: The myths of rural school and district consolidation*. (ERIC Document Reproduction No. ED135507)
- Sommer, R. (1969). *Personal space: The behavioral basis of design*. Englewood Cliffs, NJ: Prentice-Hall.
- Stevenson, K. R. (1996). Elementary school student capacity: What is the right size? *Educational Facility Planner*, 33(4), 10-14.
- Swanson, A. D. (1991, October). "Optimal" size and schooling: A relative concept. Paper presented at the annual meeting of the University Council for Educational Administration, Baltimore, MD. (ERIC Document Reproduction No. ED339139)
- University of Arkansas at Little Rock. (1997). *Factbook*. Little Rock, AR: Author, Office of Institutional Research and Budget.
- University of Michigan. (1996, April 19). Size does matter— Medium is best. *The Times Educational Supplement*, p. 19.
- Viadero, D. (1995, October). Smaller is better. *Teacher Magazine*, pp. 18-19.
- Walberg, H. J. (1989). District size and student learning. *Education and Urban Society*, 21(2), 154-63. (ERIC Document Reproduction No. EJ390005)
- Walberg, H. J. (1994). *On local control: Is bigger better?* (ERIC Document Reproduction No. ED361164)
- Walberg, H. J., & Walberg, H. J., III. (1994). Losing local control. *Educational Researcher*, 23(9), 19-26. (ERIC Document Reproduction No. EJ490373)
- Walberg, H. J., & Fowler, W. J., Jr. (1986). *Expenditure and size efficiencies of public school districts*. (ERIC Document Reproduction No. ED274471)
- Witcher, A. E. (Ed.). (1997). *Big schools, small schools: What's best for students?* Bloomington, IN: Phi Delta Kappa, Hot Topics Series.

Adolescent Problems and Coping Strategies: Alternative School Versus Non-alternative School Students

Deborah D. Vaughn, Ellen K. Slicker, and Judith Van Hein
Middle Tennessee State University

This study compared problems and coping strategies of 231 adolescents (mean age = 15 years) from rural, mid south alternative and non-alternative schools. During Phase I, students generated lists of problems that caused them stress. In Phase II, a Likert-format scale was created from which the students chose positive or negative coping strategies in reaction to problem scenarios from Phase I. The results indicated that, overall, alternative school students tended to respond to given problems by using more negative coping strategies, such as hitting, yelling, and taking drugs and alcohol, whereas non-alternative school students tended to use more positive coping strategies for the same problems, such as talking calmly to ones involved, going to family and friends for support, and participating in exercise and sports.

There is little argument that the least understood period of time during one's lifespan is that of early adolescence. Because it is also one of the most neglected phases of development, researchers are attempting to explore adolescents' problems and most stressful life events. Halstead, Johnson, and Cunningham (1993) found that participants (mean age of 14.8 years) most frequently described school, family, and social contexts as stressful situations. These findings are in agreement with the data obtained by Spirito, Stark, Grace, and Stamoulis (1991) that showed that children ages 9 - 13 years listed four common stressors: parents, siblings, school, and friends. Similarly, Stark, Spirito, Williams, and Guevremont (1989) found that school, parents, boyfriend/girlfriend, and friends were the most cited problems of students ages 14 to 17 years.

While all adolescents experience stressful life events, some cope with them better than others do. The adolescents who seem to be overburdened with problems usually have underdeveloped coping skills. Several studies indicate that those adolescents who handle stress inadequately do so as a result of dysfunctional and maladaptive behavior patterns. For example, Seiffge-Krenke (1993), investigating perceived stress and modes of coping in adolescents, found that clinical populations (adolescents with high problem intensity not yet referred

to counseling, adolescents in psychotherapy, and adolescent drug abusers) employed more ambivalent patterns of coping strategies, such as withdrawal, whereas normal adolescents (non-clinical population) most frequently chose to think out possible solutions when coping with difficulties and to use social resources, such as friends and family. Also, Spirito, Stark, and Williams (1988) found that an adolescent clinical population (pediatric patients with acute and chronic illnesses) used social withdrawal and distraction or wishful thinking more often than did healthy adolescents in attempting to cope with their problems. According to Compas, Orosan, and Grant (1993), it is during childhood and adolescence that emotion-focused coping skills (i.e., attempts to manage emotions associated with a stressor) are acquired. When emotion-focused coping skills are not developed, problem-focused coping skills (i.e., attempts to act on a stressor) are the more dominant coping skills used (Compas et al., 1993).

In contrast to the above studies that evaluated coping skills in adolescent clinical populations, the present study focuses on coping styles of a group of alternative school (AS) students. These students were placed in alternative schools due to maladaptive behaviors, such as delinquency, destruction of school property, chronic disruption of classes, chronic inability or unwillingness to cooperate with authorities, vulgar or obscene language or behavior, and sexual misconduct or harassment. These behaviors put the students at risk for becoming dropouts and delinquents. Alternative schools were instituted in order to provide a separate and more disciplined learning environment for the increasing number of at-risk students. Our hypothesis in this study is that these adolescent AS students will demonstrate more negative coping skills with a variety of stressful situations in comparison to students from public schools, (i.e., non-alternative

The authors gratefully acknowledge Dr. Stephen R. Hann for reviewing the manuscript and Jeffrey T. Bennett for assistance in data collection. Deborah D. Vaughn, B.S., is a Master's degree candidate in the Department of Psychology, Middle Tennessee State University. Ellen K. Slicker, Ph.D. and Judith Van Hein, Ph.D. are both Associate Professors in the Department of Psychology, Middle Tennessee State University. All correspondence should be addressed to: Ellen K. Slicker, Ph.D., Department of Psychology, Box X081, Middle Tennessee State University, Murfreesboro, TN 37132 or eslicker@mtsu.edu.

schools (N-AS)), who will demonstrate more appropriate coping skills.

To assess the coping abilities of these AS students, the AS and N-AS students were asked to respond to a variety of stressful scenarios. In contrast to many studies that have investigated the coping strategies of adolescents using forced-choice format (Carver, Scheier, & Weintraub, 1989; Fields & Irion, 1987; Folkman & Lazarus, 1980; Gamble, 1994; Violato & Holden, 1988), this study uses a format generated by the students themselves. Similarly, Phelps and Jarvis (1994) and Halstead et al. (1993) believe that using adult-defined, non-normative life experiences in the form of forced-choice questions does not accurately indicate the type or full range of problems adolescents experience. Therefore, the AS and N-AS students in this study were allowed to spontaneously generate a list of personal problems, which theoretically gives a better understanding of both the stressors and the coping strategies being examined.

Method

Participants

Participants in this study were from six rural, mid south public schools in two counties. Permission to conduct the study was obtained from the school superintendents. The N-AS population was chosen from two high schools and two middle schools. Entire classes from these schools were selected by the principals according to convenience of class scheduling. Two study halls were chosen from one high school, and one psychology plus one sociology class were selected from the other high school. Students from one middle school were taken from 6th, 7th and 8th grade homerooms whose teachers volunteered to participate. The second middle school allowed students from gym class to participate. The AS students were asked to participate from each of the two counties. Prior to the Phase I visit, parent/guardian consent forms were passed out to all the participants to be signed and returned. Only those students with signed parental/guardian consent forms were allowed to participate in the study. Anonymity and confidentiality of responses were assured. No teachers were allowed to be present during the study.

Demographics such as age, gender, ethnicity, and grade were gained through a cover sheet administered with each phase of the survey. All participants were predominantly middle class Caucasian students (94%). Racial proportions in both groups were similar. Phase I employed 196 students from N-AS (90 males and 106 females) and 35 students from AS (29 males and 6 females). The age range of students for Phase I was 12 to 18 years ($M = 15.0$ years, $SD = 2.6$ for non-alternative students; $M = 15.3$ years, $SD = 1.6$ for alternative students). Due to inclement weather causing school closings,

one of the middle schools was not revisited for Phase II. Also, there were four N-AS students and three AS students absent on the day Phase II was administered. Consequently, in Phase II, 68% of the students from Phase I were surveyed, 125 from N-AS (59 males and 65 females) and 32 from AS (26 males and 6 females). If students did not participate in Phase I, they were not allowed to participate in Phase II. The age range of students for Phase II was 12 to 18 years ($M = 14.9$ years, $SD = 2.3$ for non-alternative students; $M = 15.1$ years, $SD = 1.5$ for alternative students).

Procedure

During Phase I, all students from N-AS and AS were asked independently to list problems that caused them "to be upset, to worry, to cry or to act out in ways that get them in trouble." All problems from all students were grouped according to similarities. For example, if parents were the issue, then the problem fell under the category of parental problem. If the problem involved brothers or sisters, then it was categorized as a sibling problem. After all problems were categorized by one rater, inter-rater reliability was determined by having a second rater categorize 30% of the questionnaires. Between-rater agreement was 90% for categorization of problems. The categorizations of the first rater were deemed sufficiently reliable to proceed with the analysis. Problems reported in Phase I were tallied, but only those listed at least 5% of the time by at least one of the groups of students were considered further in Phase II.

From the problems listed in Phase I, a set of scenarios was developed for Phase II of the study. The materials for Phase II were patterned after Kidcope (Spirito et al., 1988). As in Kidcope, these scenarios represented the top 10 problems listed by both school types. Since parental problems and social problems (or any type of conflict with peers) were highly represented by both school types, several scenarios encompassed these problems. For example, parental arguing and longing for a divorced parent were both considered major parental problems by the students, therefore they were listed as two distinct scenarios. In addition, because of the large percentage of students listing peer problems, four different scenarios dealt with peer issues. These included being wrongfully blamed for something, being misunderstood by peers, dealing with negative peer pressure, and dealing with peer expectations. The remaining four scenarios were created based on the student-listed problems relating to boyfriends/girlfriends, siblings, self-image, and worry about others' deaths. Six coping strategies, three positive (e.g., exercising, talking calmly, and going to family or friends for support) and three negative (e.g., hitting, yelling, and using drugs or alcohol) were

ADOLESCENT PROBLEMS AND COPING STRATEGIES

listed for each scenario for a total of 60 response items. Each coping strategy was accompanied by a five-point Likert-format scale so that students could mark whether the coping strategy was never used (no), used a little, used sometimes, used usually, or was definitely used (yes).

A note was handed out and read to the students informing them that the school had a qualified counselor who was available to talk to them if generating these problems had been upsetting to them. An unexpected complimentary pen was given to all participating students. Estimated time taken to complete each phase was approximately 20 minutes.

Results

The self-generated problems were categorized by school type under 19 topics shown in Table 1. As was expected, there were commonalities between school types in regard to problems listed. Overall both AS and N-AS students listed relationship problems as their primary concern. These relationship problems encompassed problems with peers, parents, boyfriend/girlfriend, and siblings. The largest percentage of problems from both school types was with peers and parents. However, whereas N-AS students listed both peers and parents as major concerns (42% and 46%, respectively), AS students listed peers over parents as their major concern (62% and 38%, respectively). The adolescents who listed parental problems did not list many, if any, other problems, because the magnitude of the parental problems seemed to overshadow all others. Parental problems included situations involving both biological and step-parents and focused largely on difficulties with discipline or rules. Peer problems included being picked on because of size or age, difficulty getting along with a friend, worrying about a friend's behavior, arguing or fighting with peers, or just getting mad at society in general (i.e., pollution, litter, fighting, war). Boyfriend/girlfriend problems were also common with both groups of adolescents, with jealousy (i.e., worrying about losing a boyfriend/girlfriend to someone else) listed as the primary complaint.

Whereas peers, parents, and boyfriend/girlfriend problems were listed repeatedly by students from both school types, there were other problems that were more prevalent in one school type than in the other. The most striking difference between the two groups of students was in their view of school. For example, whereas none of the N-AS students listed dislike of school as a problem, 29% of the AS students listed this as a problem. The related problem of grades/homework also showed a dramatic difference between the two school populations.

Thirty-one percent of N-AS students found grades to be a problem, as compared to 3% of AS students. Worry over sports was another area of difference. This area was never mentioned as a problem by the AS students but was mentioned 5% of the time by the N-AS students. A third difference in problem areas was in regard to the police. AS students mentioned police problems 15% of the time. N-AS students listed police problems only 2% of the time.

Table 1
Percent of Students Listing Each Adolescent
Problem by School Type

Problem	School type	
	Alternative (n=35)	Non-alternative (n=196)
Peers	62	42
Parents	38	46
Dislike of school	29	0
Boyfriend/girlfriend	15	21
Police	15	2
Siblings	12	17
Self-image/expectations	9	16
Being talked about	9	8
Jobs	9	5
Money	6	9
Vehicle	6	1
Rules	6	0
Grades/homework	3	31
Others' deaths	3	12
Relatives	3	8
Temper	0	7
College	0	5
Tests	0	5
Sports	0	5

As detailed in the Methods section, in Phase II, ten scenarios were written based on problems generated in Phase I. The students were offered three positive and three negative coping strategies, representing physical, verbal, and supportive mechanisms. The students were then asked how often they used each of these coping strategies based on a five-point Likert-format scale. Tables 2 and 3 give the means and standard deviations of the positive and negative coping strategies, respectively, for each problem scenario and school type.

To evaluate the statistical differences in the responses between students in the two types of schools, a multi-variate analysis of variance (MANOVA), the accompanying univariate analyses of variance (ANOVA), and the effect sizes (d_{max} , Cohen, 1992) were computed and are shown in Table 4. For all ten scenarios, the AS students responded with significantly more negative coping strategies than did the N-AS students; however, the type of coping mechanism used differed. Whereas the

AS students were significantly more likely to respond with a negative physical response (i.e., hitting others or destroying property) for all of the problem scenarios than were the N-AS students, there was no significant difference between the two groups of students in negative verbal coping mechanisms for six of the scenarios. In addition, the AS students were significantly more likely than the N-AS students to use negative supportive coping mechanisms, such as using drugs and alcohol, in all the problem scenarios. In regard to specific scenarios, the AS students turned to drugs/alcohol significantly more often than did the N-AS students in coping with parental problems, girlfriend/boyfriend problems, as well as sibling and peer problems. Finally, when they felt wrongfully accused or had problems with self-image or peer expectations, AS students were significantly more likely to respond by physical violence than were the N-AS students. For several problems the AS students were also less likely to go to family/friends for support and were less likely to use positive physical responses (i.e., exercise) than were the N-AS students.

Table 2
Means and Standard Deviations for Positive Coping Strategies by Problem Scenario

Scenario	Positive coping strategy					
	Physical		Verbal		Support	
	A	NA	A	NA	A	NA
Parents arguing	2.81 (1.60)	2.77 (1.44)	1.94 (1.24)	2.52 (1.30)	2.10 (1.20)	2.61 (1.50)
Girlfriend/boyfriend	1.28 (0.7)	1.78 (1.29)	1.53 (1.05)	2.16 (1.32)	1.81 (0.24)	2.72 (0.14)
Wrongfully accused	1.88 (1.07)	1.87 (1.28)	2.03 (1.26)	2.94 (1.48)	2.06 (1.41)	2.90 (1.49)
Sibling(s)	1.81 (1.30)	1.51 (0.97)	2.61 (1.73)	2.78 (1.61)	1.77 (1.33)	2.14 (1.50)
Misunderstood by peers	2.09 (1.44)	2.57 (1.62)	1.97 (0.27)	2.55 (0.14)	2.13 (1.45)	2.72 (1.67)
Self-image	2.09 (1.57)	2.40 (1.56)	1.59 (1.07)	2.20 (1.45)	1.81 (1.35)	2.78 (1.60)
Missing parent	2.03 (1.43)	2.30 (1.52)	2.03 (1.51)	2.92 (1.58)	2.25 (1.57)	3.07 (1.60)
Negative peer pressure	1.78 (1.36)	2.04 (1.47)	1.88 (1.48)	2.72 (1.63)	1.78 (1.31)	2.80 (1.66)
Others' deaths	2.66 (1.84)	2.77 (1.65)	2.16 (1.53)	3.06 (1.64)	3.44 (1.63)	3.86 (1.52)
Peer expectations	2.31 (1.60)	2.33 (1.51)	2.06 (1.63)	2.85 (1.58)	1.76 (1.31)	2.79 (1.65)

Note. A = Alternative school. NA = Non-alternative school. Principal entries are means; standard deviations appear in parentheses below each mean.

Table 3
Means and Standard Deviations for Negative Coping Strategies by Problem Scenario

Scenario	Negative coping strategy					
	Physical		Verbal		Support	
	A	NA	A	NA	A	NA
Parents arguing	2.47 (1.34)	1.62 (0.96)	2.44 (1.50)	1.82 (1.13)	2.25 (1.48)	1.20 (0.55)
Girlfriend/boyfriend	3.50 (1.32)	2.59 (1.65)	3.41 (1.62)	2.91 (1.43)	2.34 (1.78)	1.40 (0.83)
Wrongfully accused	3.06 (1.72)	1.74 (1.14)	2.88 (1.66)	2.46 (1.38)	1.88 (1.50)	1.29 (0.73)
Sibling(s)	2.34 (1.56)	1.52 (0.96)	2.32 (1.42)	2.38 (1.45)	1.52 (1.06)	1.04 (0.30)
Misunderstood by peers	1.81 (1.26)	1.36 (0.83)	1.47 (0.84)	1.38 (0.87)	2.34 (1.72)	1.45 (1.06)
Self-image	3.09 (1.71)	1.78 (1.74)	3.00 (1.63)	1.98 (1.30)	2.25 (1.68)	1.31 (0.92)
Missing parent	2.16 (1.46)	1.45 (1.06)	2.12 (1.54)	1.81 (1.21)	2.06 (1.50)	1.35 (0.88)
Negative peer pressure	2.69 (1.65)	1.62 (1.05)	2.53 (1.67)	1.83 (1.78)	2.00 (1.46)	1.20 (0.67)
Others' deaths	2.84 (1.71)	1.86 (1.41)	1.72 (1.22)	1.53 (1.13)	2.88 (1.77)	1.58 (1.24)
Peer expectations	2.31 (1.57)	1.38 (0.80)	2.09 (1.44)	1.52 (0.83)	2.06 (1.63)	2.85 (1.58)

Note. A = Alternative school. NA = Non-alternative school. Principal entries are means; standard deviations appear in parentheses below each mean.

Table 4
MANOVAS, Univariate ANOVAS, and Effect Sizes for Coping Strategies by Problem

Scenario	F	df	d_{max}
Parents arguing (Wilks' lambda=.730)	9.11*	(6, 148)	
Negative physical	18.88*	(1, 154)	0.81 L
Positive physical	0.02	(1, 154)	0.03 S
Negative verbal	6.32*	(1, 154)	0.51 M
Positive verbal	5.08*	(1, 154)	0.45 M
Negative support	41.05*	(1, 154)	1.26 L
Positive support	3.19	(1, 154)	0.35 M
Girlfriend/boyfriend (Wilks' lambda=.828)	5.16*	(6, 149)	
Negative physical	8.20*	(1, 155)	0.57 M
Positive physical	4.40*	(1, 155)	0.42 M
Negative verbal	2.96	(1, 155)	0.34 S
Positive verbal	6.06*	(1, 155)	0.50 M
Negative support	21.25*	(1, 155)	0.91 L
Positive support	8.70*	(1, 155)	0.58 M

(Table continued)

ADOLESCENT PROBLEMS AND COPING STRATEGIES

Scenario	F	df	d _{acc}
Wrongfully accused (Wilks' lambda=.817)			
Negative physical	5.55*	(6, 149)	1.03 L
Positive physical	26.66*	(1, 155)	0.00 S
Negative verbal	0.00	(1, 155)	0.28 S
Positive verbal	1.94	(1, 155)	0.63 M
Negative support	9.77*	(1, 155)	0.62 M
Positive support	9.92*	(1, 155)	0.57 M
Misunderstood by peers (Wilks' lambda=.905)			
Negative physical	2.61*	(6, 149)	0.49 M
Positive physical	5.79*	(1, 155)	0.30 S
Negative verbal	2.17	(1, 155)	0.10 S
Positive verbal	0.04	(1, 155)	0.37 M
Negative support	3.08	(1, 155)	0.73 M
Positive support	11.26*	(1, 155)	0.36 M
Self-image (Wilks' lambda=.760)			
Negative physical	3.27	(1, 155)	1.02 L
Positive physical	7.86*	(1, 155)	0.20 S
Negative verbal	26.75*	(1, 155)	0.74 L
Positive verbal	1.05	(1, 155)	0.43 M
Negative support	15.10*	(1, 155)	0.68 L
Positive support	4.90*	(1, 155)	0.62 M
Missing parent (Wilks' lambda=.905)			
Negative physical	0.86*	(6, 149)	0.52 M
Positive physical	6.64*	(1, 155)	0.18 S
Negative verbal	0.68	(1, 155)	0.32 S
Positive verbal	2.54	(1, 155)	0.57 M
Negative support	7.88*	(1, 155)	0.68 L
Positive support	12.09*	(1, 155)	0.51 M
Negative peer pressure (Wilks' lambda=.791)			
Negative physical	6.59*	(6, 149)	0.89 L
Positive physical	20.38*	(1, 155)	0.18 S
Negative verbal	0.82	(1, 155)	0.54 M
Positive verbal	7.50*	(1, 155)	0.53 M
Negative support	7.11*	(1, 155)	0.90 L
Positive support	20.72*	(1, 155)	0.64 M
Others' deaths (Wilks' lambda=.835)			
Negative physical	10.36*	(1, 155)	0.67 L
Positive physical	4.91*	(1, 155)	0.06 S
Negative verbal	11.19*	(1, 155)	0.17 S
Positive verbal	0.11	(1, 155)	0.56 M
Negative support	0.67	(1, 155)	0.95 L
Positive support	8.00*	(1, 155)	0.27 S
Peer expectations (Wilks' lambda=.832)			
Negative physical	2.12	(1, 155)	0.93 L
Positive physical	6.09*	(5, 151)	0.01 S
Negative verbal	22.17*	(1, 155)	0.58 M
Positive verbal	0.00	(1, 155)	0.50 M
Negative support	8.67*	(1, 155)	0.74 L
Positive support	6.25*	(1, 155)	0.38 M

Note. Effect size designations based on Cohen (1992). L = Large. M = Medium. S = Small.
*p ≤ .05.

Discussion

In the open-ended format of Phase I of the study both N-AS and AS students generated problems common to most adolescents. For example, both groups repeatedly listed peer, parental, boyfriend/girlfriend, and sibling problems. Unexpectedly however, many problems differed between the two groups of adolescents (i.e., school dislike, police, grades, college, tests, and sports). Whereas 29% of AS students view school as a problem or stressor, none of the N-AS students indicate that this is a concern. Interestingly, in contrast to being listed by only a few AS students, grades and homework are listed as a problem by almost a third of the N-AS students. Perhaps the AS students, who dislike school in general, find grades and homework unimportant and thus not a major concern. This could also explain why the AS students do not list any problems concerning tests or college entrance.

Another problem area that is not listed by the AS students is sports. Either these students are not allowed to participate in school sports or their dislike of school has extended into after-school sports as well. Finally, a difference appears between the two groups regarding police. Few N-AS students are concerned with police, whereas a significant percentage of the AS listed police as a problem. This appears to be related to overall problems with rules, which is also listed by some of the AS students, but none of the N-AS students. Trouble with police and rules could account, at least in part, for behaviors that cause some of the adolescents to be placed in an alternative school in the first place.

The problems listed by both groups of adolescents (i.e., parents, society, and boyfriend or girlfriend relationships) are consistent with previous studies that have used the open-ended question format. Phelps and Jarvis (1994) asked high school students to report a specific important problem that they had dealt with in the past two months, including what happened, where it happened, who was involved, and what made the event important. They found that conflicts with parents, peers, and boyfriends ranked the highest for girls whereas concerns about grades and extracurricular activities ranked highest for boys. Brown, O'Keefe, Sanders, and Baker (1986) asked adolescents to spontaneously generate personal problems in their lives. The most frequently cited problem was fear of negative evaluation, followed by a fight with/or rejection by the opposite sex, conflict with an adult, and lastly, concerns regarding the future.

In contrast to the present study and that of Brown et al. (1986), Violato and Holden (1988) asked adolescents to complete a questionnaire consisting of adult-defined,

forced-choice questions. Violato and Holden concluded that the most representative concerns of adolescents are future and career, followed by health and drugs, personal self, and social self. These concerns are clearly inconsistent with those generated by the adolescents themselves in this study and the study by Brown et al. (1986). Similarly, Adwere-Boamah and Curtis (1993) also offered adult-defined problems to a lower socioeconomic-status (SES) group of adolescents and similarly found career issues, grade issues, future schooling issues, and appearance issues were listed as the four most serious concerns from the response options offered. Once again, these adult-generated choices do not appear to encompass those issues that are actually of concern to adolescents.

Phase II addresses the coping strategies used by N-AS and AS students for the same set of problems. As was expected, the AS students use negative coping strategies significantly more often than do the N-AS students, whereas the opposite is true for positive coping strategies, even when considering the same problems. For example, when AS students are confronted with a boyfriend/girlfriend relationship problem, their strategies involve physical violence and the use of drugs or alcohol as responses. When N-AS students are asked to respond to the same problem, they most often select positive strategies (i.e., talking calmly to others involved, seeking family/friend support).

Responding through negative physical behavior and negative support mechanisms, such as using drugs and alcohol, seems to be the typical pattern for these AS students. It appears that their negative coping choices actually contribute to the likelihood of their being placed in an alternative school. Much of the past research in the area of adolescent coping has focused on problem-focused coping (e.g., attempts to act on a stressor) and emotion-focused coping (e.g., attempts to manage one's emotions associated with a stressor) in relation to one's age and cognitive development (Compas et al., 1993). This study differs from previous research in two ways. First, it compares coping strategies of both AS and N-AS students, finding that AS students use a greater percentage of negative coping strategies than do N-AS students on a matched set of common problems. Second, it looks at specific behaviors that are used to cope with specific stressful situations.

One of the limitations of this study is the lack of random selection of participants from the various schools. In addition, these schools represent only a small geographical area, so caution should be exercised when generalizing these results to other populations. Finally, the measure used to assess coping strategies was developed specifically for this study; therefore, additional

research is necessary to establish the measurement reliability of this scale.

The present study may provide a better understanding of the problems and coping strategies used by at-risk students. It may also be useful for identifying students with negative, self-defeating coping strategies. Perhaps if these students were identified early enough, they could be taught more appropriate coping methods before maladaptive behavior occurs, thus providing a smoother progression through their adolescent years and schooling.

References

- Adwere-Boamah, J., & Curtis, D. A. (1993). A confirmatory factor analysis of a four-factor model of adolescent concerns revisited. *Journal of Youth and Adolescence, 22*, 297.
- Brown, J. M., O'Keeffe, J., Sanders, S. H., & Baker, B. (1986). Developmental changes in children's cognition to stressful and painful situations. *Journal of Pediatric Psychology, 11*, 343-357.
- Carver, C., Scheier, M., & Weintraub, J. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality and Social Psychology, 56*(2), 267-283.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin, 112* (1), 155-159.
- Compas, B. E., Orosan, P. G., & Grant, K. E. (1993). Adolescent stress and coping: Implications for psychopathology during adolescence. *Journal of Adolescence, 16*, 331-349.
- Fields, F. B., & Irion, J. C. (1987). Coping strategies from the perspective of two developmental markers: Age and social reasoning. *Journal of Genetic Psychology, 149*(2), 141-151.
- Folkman, S., & Lazarus, R. S. (1980). An analysis of coping in a middle-aged community sample. *Journal of Health and Social Behavior, 21*, 219-239.
- Gamble, W. C. (1994). Perceptions of controllability and other stressor event characteristics as determinants of coping among young adolescents and young adults. *Journal of Youth and Adolescence, 23*, 65-84.
- Halstead, M., Johnson, S., & Cunningham, W. (1993). Measuring coping in adolescents: An application of the ways of coping checklist. *Journal of Clinical-Child-Psychology, 22*, 337-334.
- Phelps, S. B., & Jarvis, P. A. (1994). Coping in Adolescence: Empirical evidence for a theoretically based approach to assessing coping. *Journal of Youth and Adolescence, 23*, 359-371.
- Seiffge-Krenke, I. (1993). Coping behavior in normal and clinical samples: more similarities than differences? *Journal of Adolescence, 16*, 285-303.

ADOLESCENT PROBLEMS AND COPING STRATEGIES

- Spirito, A., Stark, L. J., & Williams, C. (1988). Development of a brief checklist to assess coping in pediatric populations. *Journal of Pediatric Psychology*, 13, 555-574.
- Spirito, A., Stark, L. J., Grace, N., & Stamoulis, D. (1991). Common problems and coping strategies reported in childhood and early adolescence. *Journal of Youth and Adolescence*, 20, 531-544.
- Stark, L., Spirito, A., Williams, C., & Guevremont, D. (1989). Common problems and coping strategies: I. Findings with normal adolescents. *Journal of Abnormal Child Psychology*, 15, 203-212.
- Violato, C., & Holden, W. B. (1988). A confirmatory factor analysis of a four-factor model of adolescent concerns. *Journal of Youth and Adolescence*, 17, 101-113.

Appendix

Problem Scenarios

Scenario A: Your parents have been arguing and hollering at each other for the past 30 minutes.

Scenario B: You are at a party and you see your girlfriend or boyfriend kiss someone else.

Scenario C: An adult is accusing you of something that you did not do. Because you cannot prove that you didn't do it, you will be punished. For example, you might get grounded, have your car taken away, get expelled from school, or a teacher might give you an F.

Scenario D: Your brother or sister has taken your favorite jacket without asking and left it at a friend's house.

Scenario E: You feel like you don't have any friends and no one understands you at all.

Scenario F: You do not like your looks (for example, too fat, too skinny, too tall, too short, nose too big, nose too small, face has lots of zits, ears too big, etc). You just heard someone make fun of your looks.

Scenario G: Your parents are divorced. You really miss the parent that you don't live with because they never call or come around, not even on your birthday.

Scenario H: Some of your friends are pressuring you to do something that you do not want to do, they call you names, and they tell you that they won't have anything to do with you anymore unless you do it.

Scenario I: Someone you love has died.

Scenario J: You feel like everyone is pressuring you to be someone better than who you are, or they are pressuring you to be someone different than who you are.

Coping Strategies

How likely are you to deal with the problem by:

1. Hitting something, hitting somebody, or destroying property.
2. Exercise, ride a bike, play basketball, go jogging, workout, etc.
3. Threaten to get even, yell, scream, argue, or blame the other person(s).
4. Talk calmly to the other(s) involved in order to reach an understanding.
5. Use drugs or alcohol as a way of escaping the problem.
6. Go to family or friends for support and comfort.

High-Stakes Accountability and Contextual Effects: An Empirical Study of the Fairness Issue

Edward B. Reeves
Morehead State University

Are high-stakes accountability measures fair to all school systems despite disparities of wealth, community mores, and geographic location? This issue is addressed using Kentucky school district and grade-level accountability data. A repeated measures general linear model approach is employed to test two ways in which contextual effects can influence accountability measures: (a) gains in within-district scores and (b) differences in scores achieved by different districts. The findings show that contextual effects have little to no influence on within-district gains. However, contextual effects do account for 30-40% of the between-districts differences in scores. Median household income is the strongest of these effects, followed by rural-metro differences and teen birth rate. The contextual effects increase with grade level. The results of this study help to alleviate concern about bias when within-district gains are used for deciding accountability. However, comparisons between districts using unadjusted scores risk substantial bias because of contextual effects.

Recent educational reform initiatives in the United States have focused on monitoring and improving outcomes and holding school districts and schools accountable for what their students learn. Currently, there are two prevailing views about how to obtain greater accountability. The first view opts for creating greater choice of educational services through magnet schools, voucher plans, charter schools, and tax rebates that offset the costs of alternative schooling. According to this view, parents in the role of educational consumers should be free to choose the school they believe will best educate their child. This perspective relies on market competition among schools (within the public sector but also between the public and private sectors) to foster academic excellence and accountability. The second view relies on an administered system of accountability to recognize and reward success, specifically in the public schools. This second approach has the advantage of the first since it keeps the present administrative structure of public school systems largely intact while reforming that structure to achieve accountability goals. Realistically speaking, an administered accountability system may be the only choice for states like Kentucky that have a large number of rural school districts. The higher cost of delivering education to rural areas means that parents have fewer

educational choices available (Ladd, 1996; see also Federal Reserve Bank of New York, 1998).

Prior to the 1990's, Kentucky students had a record of performing below the national average on achievement tests, and high school graduation rates were low. The Kentucky Education Reform Act of 1990 (KERA) launched one of the nation's most ambitious educational reform efforts of recent decades (Elmore, Abelman, & Fuhrman, 1996). Among KERA's most noteworthy accomplishments has been the high-stakes performance assessment program, the Kentucky Instructional Results Information System (KIRIS). This program embraced a complex procedure for testing students using authentic assessment methods. The KIRIS test results have become the basis for granting monetary rewards to schools that showed significant improvement and for levying sanctions against those that failed to show progress (Guskey, 1994). Controversial from the start, KIRIS has been attacked persistently throughout the 1990's. Recently, a panel of external evaluators has recommended numerous changes in the assessment tests and procedures (Catterall, Mehrens, Ryan, Flores, & Ruben, 1998). The accountability system is being overhauled during the 1998-99 school year. Despite the controversy, the demand in Kentucky for an effective means of assessing and improving educational performance has not abated. Whatever its shortcomings, KIRIS put educators on notice that improved performance is expected.

The Present Study

The present study deals with two closely related questions: Are high-stakes accountability measures fair to all school systems despite disparities of wealth,

I would like to thank Dan Fasko, Debbie Grubb, Bud Harty, Ben Oldham, Peggy Reeves, Dave Rudy, and three anonymous referees for thoughtful comments on several drafts of this paper. Correspondence concerning this article should be addressed to Edward B. Reeves, R & D Center for Students, Schools, and Communities, UPO Box 785, Morehead State University, Morehead, KY 40351 or via e-mail at e.reeves@morehead-st.edu.

community mores, and geographic location? What method can be used to find this out? In order to address these questions, the Kentucky accountability index scores for public school districts during two biennia, 1992-94 and 1994-96, are examined. The purpose of the analysis is to find out the extent to which the scores are associated with contextual effects and thus to ferret out possible unfairness in Kentucky's high-stakes accountability system.

Research on Contextual Effects and Educational Outcomes

In the present study, "contextual effects" refer to factors in the school district's social environment that influence aggregated student performance on the KIRIS accountability test. Contextual effects are defined to be external to the schools and not subject to the direct control of teachers and administrators. Studies showing that contextual effects influence educational outcomes have become a staple in the literature. *Equality of Educational Opportunity* (Coleman et al., 1966) reached the (at the time surprising and highly controversial) conclusion that much variation in student performance was not so much a consequence of the differences in school inputs and resources as the disparate family and community backgrounds of students. Since this watershed study, many researchers (e.g., Coleman, 1988, 1990; Hallinan, 1988; Schneider & Coleman, 1993; Smith, Beaulieu, & Israel, 1992; Teachman, Paasch, & Carver, 1997) have reaffirmed the importance of contextual effects. These findings have become so commonplace and well understood that effective schools research commonly employs family socioeconomic indicators to capture external factors that influence learning (Hannaway & Talbert, 1993; Hanushek, 1997; Lee, Smith, & Croninger, 1997; Zigarelli, 1996). In Kentucky, one pre-KERA study noted the effect of economic deprivation on school district performance (Guskey & Kifer, 1990). Similar findings were reported more recently in Alabama (Mathews & Hackett, 1997). To date, most of the research has focused on how student achievement varies with the family's socioeconomic circumstances and social capital resources.¹ Community-level effects on educational outcomes have been studied less, although examples of this type of research do exist (Crane, 1991; Garner & Raudenbush, 1991).

The objective of the present study is to contribute to the research-based literature that examines contextual effects and educational outcomes. For reasons discussed below, this study focuses on how community context affects school district accountability index scores. If appropriate caution is exercised, the findings discussed below may be extrapolated to schools.

Method

Simple cross-tabulations and correlations are inadequate to address the problem of contextual effects and the possibility of bias in accountability scores. Accountability outcomes have multiple causes (Thompson, 1992), and some of these causes may not be understood by the designers of the accountability system (Catterall et al., 1998). A multivariate approach is necessary to sort out the causal links. The analysis employed here uses repeated measures general linear modeling (GLM), a procedure that combines multivariate ANOVA with multiple regression, to test for two distinct sources of bias in Kentucky's accountability system: (a) if a gain in accountability scores within school districts is attributable to any of several contextual effect variables and (b) if between-district scores are influenced by the same contextual effects. Since repeated measures analysis is sensitive to departures from normality, Box's test of equality of covariance matrices and Levene's test of equality of error variances were applied to the data as a check on the assumptions (Keselman et al., 1998). For all analyses described below, the null hypotheses of Box's and Levene's tests were rejected after the distributions of the dependent variables were log transformed.

Data

The Kentucky Department of Education has made the annual accountability index scores for each school and school district publicly available since the 1992-93 school year. The scores are computed in part from tests administered to fourth, eighth, and twelfth graders² and in part from noncognitive criteria, including attendance, retention, dropout, and transition to adult life. The components of the accountability index are as follows: reading, mathematics, science, social studies (each weighted 14% of the total score), arts & humanities, practical living (each weighted 7%), on-demand writing (weighted 3.5%), portfolio writing (weighted 10.5%), and noncognitive data (weighted 16%).

KIRIS testing has been guided by an educational philosophy that stresses "authentic learning" and effective writing rather than rote memorization of discrete facts. Thus, the mode of testing relies heavily on student portfolio development and open-response questions. One consequence of this philosophy is that the usual procedures for determining the validity and reliability of test questions do not have a ready application to KIRIS. Sponsors of the test have argued that the influence of high-stakes open-response and portfolio assessment on instruction is sufficiently beneficial to outweigh the disadvantages (see Catterall et al., 1998, for a meticulous

discussion of validity and reliability issues with respect to KIRIS).

Selection of Variables

In Kentucky, greater attention is devoted to school performance than to district performance. Therefore, the most germane point of attack for the present study would have been to investigate contextual effects on school performance. Unfortunately, good measures of school-level contextual effects were not available. Several measures often used to suggest contextual effects would be better used as measures of student body characteristics -- for example, the percent of students receiving free and reduced-cost lunch or the percent of minority students enrolled in the school. These measures would not accurately convey the composition of the communities in which the schools are found. Characteristics of the schools themselves, such as per student spending and teacher-student ratio, would have been even less satisfactory measures of contextual effects. In contrast with the unsatisfactory state of school-level context measures, good measures of contextual effects associated with school districts were found in census reports and similar sources. These data were chosen for the present analysis because they accurately characterized the local population within the school district.

Dependent Variables

School District Accountability Index Scores. In Kentucky, the margin of improvement in student scores during one two-year period compared with the preceding period has been the basis for determining accountability. Biennial weighted mean scores have been used because they have proven to be more reliable than single-year mean scores. Thus, one set of dependent variables used in this study consists of the 1992-94 and 1994-96 weighted mean scores for each school district.

Grade-Level Accountability Index Scores. A second set of dependent variables consists of grade-level scores aggregated by school district. These scores are used to test for corroboration with the overall district-level results. In addition, analysis of the grade-level scores provides insight into the contextual effects at a level of aggregation more closely approximating the schools. Because weighted biennial scores were unavailable for the grade-level analysis, two-year simple mean scores for each grade level were calculated from the annual scores which the Kentucky Department of Education provided. For example, a 1992-94 fourth-grade district mean score was calculated by averaging the 1992-93 and 1993-94 fourth-grade mean scores. Since the resulting biennial mean is unweighted by grade enrollment, measurement error may occur. However, the assumption is made that district-

level grade enrollments do not change sufficiently from the one academic year to the next to invalidate this estimate of the biennial grade-level performance.

Independent Variables

In looking for variables to represent contextual effects, the goal was to choose a mix of factors, including an SES measure, such as household income, and also geographic and demographic characteristics of the county where the school district is located. A measure of the strength of local norms that encourage commitment to schooling was sought as well.

Rural-metro differences--a fixed effect. A "rural-metro index" was constructed to measure the combined effect of geographic location and population concentration on the KIRIS results. Although some researchers suggest that nonmetro and metro student achievement has equalized in recent years (e.g., Greenberg & Teixeira, 1995), most believe that geographic location and demographic characteristics continue to exert an important influence on schooling. Still, the advantages of metropolitan versus rural education are not clear-cut and one-sided. The concentration of population in metro areas offers greater amenities and resources than can be mustered in small towns and rural districts, and, other things being equal, this makes the delivery of educational programs and services less costly (Stern, 1994). Rural schools may also experience comparative disadvantages resulting from a narrow occupational structure in which career opportunities depend on family ties and localized social networks. In these circumstances students and parents are less likely to see the connection between doing well in school and career opportunities (Marino, 1995). Rural power structures may pose greater obstacles to progressive educational efforts (DeYoung, 1991). Offsetting the adverse effects of a rural locale, rural communities can sometimes offer advantages that are less available in the metropolis. Social capital can be more prevalent and school-community relations can be more close-knit in rural than in metropolitan areas (Khattri, Riley, & Kane 1997). Furthermore, the negative effect that low family SES has on student performance may be mitigated by the greater mixing together of students from divergent SES backgrounds in rural classrooms (Alspaugh, 1992).

The rural-metro index was constructed by recoding the USDA Economic Research Services' 1993 Urban Influence Codes. The Urban Influence Codes classify all U.S. counties into nine categories based on the size of the Metropolitan Statistical Area (MSA) for metro counties, adjacency to MSA's, and size of the largest city for non-metro counties. To create the index used for the present study, the first two categories of the Urban Influence

Codes, which differentiated small from large metropolitan areas, were collapsed. The next four categories were also collapsed into one metro adjacent category. The remaining categories of the Urban Influence Codes -- rural counties differentiated by the size of the largest city -- were left as is. Thus, the index used in this study has five categories, coded 1 to 5 (see Table 1). This recoding of the Urban Influence Codes was prompted by the fact that Kentucky counties are not well represented in certain metro and metro adjacent categories. In the GLM analysis, the rural-metro index is treated as a fixed effect, since the relationship between the rural-metro index and the accountability scores is nonlinear, as will be seen.

Table 1
The Rural-metro Index

Code	Description
1	Nonmetro counties not adjacent to a metro area and with no city or a city with a population less than 2,500
2	Nonmetro counties not adjacent to a metro area and with a city of 2,500 to 9,999 population
3	Nonmetro counties not adjacent to a metro area and with a city of 10,000 or more population
4	Nonmetro counties physically adjacent to an MSA with at least 2 percent of the employed labor force in the county commuting to the metro area
5	Metro counties (i.e., located in an MSA)

Note. Adapted from the 1993 Urban Influence Codes, Economic Research Service, U.S. Department of Agriculture.

Other contextual effects. After some alternatives were considered, median household income in the school district was chosen to measure the socioeconomic context.³ This variable was taken from the School District Data Book (National Center for Educational Statistics), which is based in turn on the 1990 U.S. Census. Median household income was discovered to be positively related to the scale of the rural-metro index codes (see Table 2).

Another contextual effect was the teen birth rate, the number of births in each county per 1,000 females, aged 12 to 17 years, in 1992-94. This variable was taken from the 1995 Kentucky KIDS COUNT and was chosen as a proxy for community norms that negatively influence educational achievement and persistence. A high teen birth rate suggests a community that has limited occupational choices for young adults, higher levels of family instability, and low orientation toward secondary school completion and postsecondary education (Bickel, Weaver, Williams, & Lange, 1997; Rodriguez & Moore, 1995). In

the present study, teen birth rate is inversely related to both the rural-metro index scale and median household income (Table 2).

Table 2
Characteristics of School Districts by Rural-metro Index Code

Rural-metro index	# of school districts	Median household income	Teen birth rate	# of independent school districts
1 Rural, < 2,500	30	\$15,026	22.8	2
2 Rural, 2,500-9,999	41	\$16,952	23.5	14
3 Rural, ≥ 10,000	19	\$18,678	22.2	11
4 Metro adjacent	44	\$20,577	18.9	10
5 Metro	42	\$27,602	18.5	19
Total	176	\$20,258	20.9	56

Note. Median household income and teen birth rate express the average for each rural-metro category and for the total sample.

A control variable. The last variable, a dummy, indicated if the school district is independent. Besides the 120 county school districts in Kentucky, there are 56 independent school districts, enclaved within 40 of the counties. The most independent school districts that any county has is six, while three-quarters of the forty counties have only one independent district apiece. In a preliminary analysis it was found that the independent districts outperform the county districts and that independent districts are most common in metro counties and least common in rural counties having a city with a population less than 2,500 (see Table 2). Being smaller and more geographically compact than the county school districts, independent districts may have reaped administrative efficiencies from less bureaucracy, lower transportation costs, and a stronger organizational culture. Another advantage of independent districts could be that they have stronger support from their local communities than county school districts. Since the presence of independent districts could have confounded the tests of the contextual effects, this variable entered the analysis as a control.

Research Design and Hypotheses

Contextual effects can influence district performance measures in two basic ways. First, the effects can influence progress or the lack of it within each school district. That is to say, because of adverse contextual effects some districts could have their ability to improve inhibited. Other districts because of beneficial contextual effects could receive a boost in performance. The discovery of a finding like this would be a blow to Kentucky's educational reform effort. One of KERA's guiding principles

HIGH-STAKES ACCOUNTABILITY AND CONTEXTUAL EFFECTS

is that all schools and school districts (and all students) can significantly improve their performance, no matter what their background of advantages or disadvantages. A second way to show the consequences of contextual effects for accountability scores is with a between-districts comparison. Districts with adverse contexts could be found to score lower than districts with beneficial contexts, although the lower scoring districts might register improvement when compared with their own previous effort. Since Kentucky allocates rewards and sanctions according to how well a district or school performs compared with its previous benchmark, between-districts comparisons are of less concern for the state's accountability initiative. Nevertheless, the news media uses the accountability index scores to compare and rank districts as well as schools. The discovery that contextual effects bias such comparisons could serve to inform the media and the public of the danger of simplistic comparisons.

Two hypotheses are tested with respect to the overall district and the grade-level accountability scores:

1. Contextual effects influence the within-district change in accountability scores between the 1992-94 and 1994-96 biennia.
2. Contextual effects influence the between-district differences in accountability scores in each biennium.

Analysis and Results

Repeated Measures General Linear Model

Repeated measures GLM is an appropriate method when measurement is made more than once on each subject or case. The procedure produces two distinct types of models. In the first type, the factor labeled WDS (within-district scores) represents the 1992-94 to 1994-96 change in scores for each district. The within-district model then tests for differences between each district's two-year mean scores and uses interaction terms to estimate the effect of the contextual variables on the within-district change in scores. The contrasting between-districts model divides the sample of subjects or cases into discrete subgroups related to the fixed effect, which in the present study is the rural-metro index. The analysis of the between-districts effects estimates regression parameters for the fixed effect and the other independent variables.

Analysis of District Scores

Within-District Model. Between the biennia of 1992-94 and 1994-96 the overall mean accountability score for the entire state increased from 39.3 to 44.6 (an increase of 13.5%). If this study were interested only in finding out if the within-district scores had significantly increased,

repeated measures GLM would be unnecessary. A paired sample *t*-test to compare the difference between the means would suffice. However, since we wish to know what differences (if any) the contextual effects make, we test their interactions with the within-district scores. As Table 3 shows, the test of the within-district differences in scores is significant and, as indicated by *eta* squared (η^2), the size of effect is moderate (18.9% of the total variation). Turning to the interaction terms, median household income is the only contextual effect to make a statistically significant contribution to the variation in the within-district scores. The *eta* squared value shows that the size of this effect is relatively small, because it explained only 3.3% of the total variation. Contextual effects, therefore, have little to no impact on the gain in the within-district scores.⁴

Table 3
Tests of Within-district Effects

Source of the effect	df	F	η^2
Within-district scores (WDS)	1	39.090***	.189
WDS x Med HH income	1	5.817*	.033
WDS x Teen birth rate	1	.861	.005
WDS x Independent district	1	1.659	.010
WDS x Rural-metro index	4	1.257	.029
Error (WDS)	168		

* $p < .05$. *** $p < .001$

Between-Districts Model. The parameter estimates for the between-districts model (see Table 4) show that, after controlling for the influence of independent districts, most of the contextual measures are statistically significant. Median household income is a strong positive predictor of higher scores in both biennia, while teen birth rate has a negative influence. In 1992-94, the effects of the first four categories of the rural-metro index are significant and positive. Since the reference category consists of metro counties, this means that metro school districts are being outperformed by all other rural-metro categories. In 1994-96, rural counties whose cities have populations greater than 2,500 significantly outperform all other categories, while there are no significant differences between metro, metro adjacent, and the most rural school districts.

A multivariate test of the between-districts effects that averages the results from both biennia yields the following *eta* squared values: median household income (0.244), teen birth rate (0.063), independent district (0.067), and rural-metro index (0.090). Thus, across both biennia, rural-metro differences rank second in importance after median household income.

Table 4
Parameter Estimates of Between-districts Effects

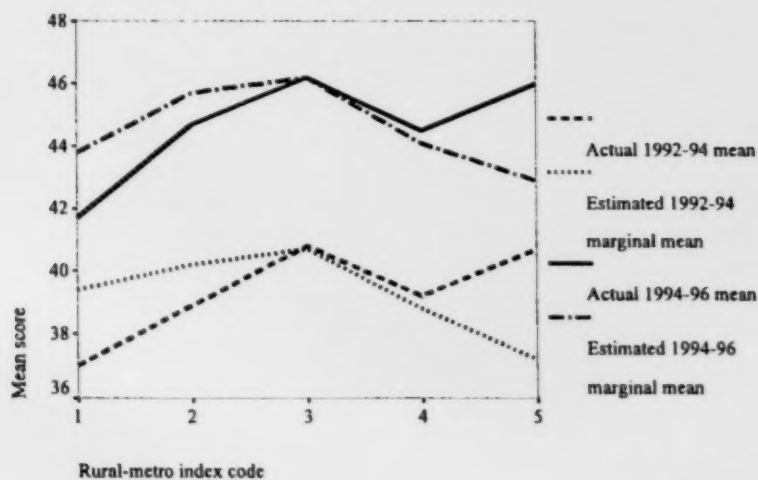
Dependent variable	Parameter	B	SE B	η^2
1992-94 district score (ln) ^a				
	Intercept	3.510***	.042	
	Med HH income	8.421E-06***	.000	.271
	Teen birth rate	-3.856E-03***	.001	.070
	Independent district	5.442E-02	.014	.079
	Rural-metro index			
	1 Rural, < 2,500	5.576E-02*	.025	.029
	2 Rural, 2,500-9,999	7.669E-02**	.022	.069
	3 Rural, $\geq 10,000$	9.012E-02***	.025	.073
	4 Metro adjacent	4.132E-02*	.020	.025
	5 Metro	0 ^c		
1994-96 district score (ln) ^b				
	Intercept	3.677***	.044	
	Med HH income	6.683E-06***	.000	.178
	Teen birth rate	-3.176E-03**	.001	.045
	Independent district	4.190E-02**	.015	.045
	Rural-metro index			
	1 Rural, < 2,500	1.949E-02	.026	.003
	2 Rural, 2,500-9,999	6.330E-02**	.023	.045
	3 Rural, $\geq 10,000$	7.229E-02**	.026	.045
	4 Metro adjacent	2.650E-02	.021	.010
	5 Metro	0 ^c		

^aAdjusted $R^2 = .423$. ^bAdjusted $R^2 = .331$. ^cThis parameter is set to zero because it is redundant.

* $p < .05$. ** $p < .01$. *** $p < .001$.

How well the between-districts analysis predicts the scores can be judged from the adjusted R-squared coefficients. The 1992-94 model explains 42% of the variance in scores. In the next biennium, the explained variance drops to 33%. Furthermore, the *eta squared* values reveal that the decline in the effects of the contextual variables across the two biennia occur in all variables. Whether this decline in the strength of the contextual effects portends an important trend cannot be decided from only two biennia of performance data.

A graphical presentation illustrates more clearly how the between-districts comparisons relate to rural-metro differences and the nonlinearity of these effects (see Figure 1). It reveals that when the effects of median household income, teen birth rate, and independent school districts are held constant, the more rural school districts perform better than their actual mean scores suggest, while metro districts perform worse. The graph also shows that the scores of the rural counties with the largest towns and the metro adjacent counties are least affected by median household income and teen birth rate. Rural counties with towns larger than 10,000 in population appear to be the optimal environment for educational performance in Kentucky.



Note. Marginal means are estimated from GLM analysis in Table 4.

Figure 1. Actual Means v. Estimated Marginal Means of Accountability Scores by Rural-metro Index

HIGH-STAKES ACCOUNTABILITY AND CONTEXTUAL EFFECTS

Analysis of Grade-Level Scores

This study focuses on district accountability scores because good measures of contextual effects are available at this level of aggregation. These results cannot be extrapolated to the school level without danger of aggregation bias. Nevertheless, since mean scores for grades 4, 8, and 12 by district are available, it is possible to test for the presence of contextual effects at different grade levels within and between districts. While not a school-level analysis as such, this is a closer approximation than in the previous district-level tests. The grade-level analyses will also reveal if the contextual effects vary by grade. The findings are summarized without a tabular presentation of the full results of each model.

Grade 4. The mean accountability score of fourth graders increases from 38.3 in 1992-94 to 46.5 in 1994-96 (a gain of 21.4%). The test of the within-district scores confirms that the improvement is statistically significant; however, the contextual effects make no contribution to the explained variance. Weak contextual effects are detected in the between-districts model, although median household income alone is statistically significant in both biennia. As in the district-level analysis, the adjusted R-squared coefficients of the fourth-grade model show a decrease between the first and second biennium (see Table 5).

districts and rural districts with cities of 2,500 or more inhabitants. In the eighth grade, the R-squared coefficients of the between-districts model are larger than in the fourth-grade models, but they also reveal a decline across the two biennia (Table 5).

Grade 12. The twelfth-grade mean accountability score shows an increase from 39.5 in 1992-94 to 44.1 in 1994-96 (a gain of 11.6%). A paired samples *t*-test finds this change significant ($t = 15.099$, $df = 170$, $p < .001$). However, the change is no longer significant in the multivariate model with the inclusion of median household income and teen birth rate together, although neither of these variables is significant. Neither the rural-metro index nor independent district significantly influence the within-district analysis of variance.

The between-districts effects are the most pronounced at the twelfth-grade level, with results that closely approximate the overall district results. Median household income is the strongest contextual effect. Twelfth graders in metro and metro adjacent school districts perform less well than in rural districts, and the influence of teen birth rate is negative. The adjusted R-squared coefficients for the twelfth grade model are larger than for either the eighth or fourth grade. A decline between the first and second biennium repeats a pattern already established by the preceding analyses (Table 5).

Table 5
Adjusted R-squared Coefficients for the Between-districts
Model by Grade Level and Biennium^a

	1992-94	1994-96
Grade 4	.156	.086
Grade 8	.318	.253
Grade 12	.417	.335
Overall	.423	.331

^aThe general between-districts model is: Accountability index score (ln) = Intercept + b_1 (Med. HH income) + b_2 (Teen birth rate) + b_3 (Independent district) + b_4 (Rural, < 2,500) + b_5 (Rural, 2,500-9,999) + b_6 (Rural, $\geq 10,000$) + b_7 (Metro adjacent) + b_8 (Metro).

Grade 8. This grade level registers an increase in the mean score from 40.0 to 43.1 (a gain of 7.8%). Although this gain is less than the fourth-grade increase, it too is statistically significant. Again, the contextual effects do not contribute to the within-district gain. Between the districts, however, median household income, teen birth rate, and rural-metro differences prove to be statistically significant effects for the eighth grade, with median household income having the strongest effect in both biennia. Metro school districts fare less well on the KIRIS tests than nonmetro categories at this grade level. The largest differences in scores are between metro

Discussion

Contextual effects have been shown to have little impact on the change in within-district scores. The main exception was found in the twelfth grade where median household income and teen birth rate together washed out the apparent increase in the scores. A different story must be told for between-districts comparisons, however, because 30-40% of the variation in scores is accounted for by contextual effects. Clearly, district-to-district comparisons cannot be fairly made without first adjusting for these effects. Median household income has the strongest influence on between-districts comparisons, followed by rural-metro differences and teen birth rate. After holding median household income, teen birth rate, and independent school district constant, rural school districts generally outperform metro districts. The highest scores are found in rural counties where the city has a population of 10,000 or more. The next highest scores are in rural counties with cities having a population ranging from 2,500 to 9,999. But even in the most rural counties (those with cities having populations less than 2,500), school districts perform about as well as metro districts and sometimes better. No detailed explanation of this finding is possible using the data available. However, one may propose that after the disadvantages of low household

income and higher teen birth rate are partialled out, rural school districts have comparative advantages owing to greater social capital and stronger school-community relations than is typical of metropolitan school districts (Khattri et al., 1997; Stern, 1994).

The grade-level analyses generally corroborate the overall district-level analysis. Moreover, since the grade-level analyses approximate a school-level analysis, these results offer some basis for a cautious extrapolation of the findings to schools. The findings suggest that contextual effects have no influence on school performance gains, except possibly at the high school level. Lower grades are less influenced by contextual effects than higher grades, a finding that is reminiscent of the Coleman Report (Coleman et al., 1966) and one that suggests contextual effects are more pronounced in older student populations.

The present findings also speak to an issue debated in a prominent Kentucky newspaper. On December 10, 1997, a front-page article in the *Lexington Herald-Leader* reported on a proposal by the state's Office of Education Accountability (OEA). The proposal called for changing the way accountability scores would be evaluated. To increase the fairness of the accountability system, the OEA recommended that school scores be adjusted for student poverty rates, education level of parents, per pupil spending, class and school size, teacher salaries, and the district's tax rate. In an editorial The *Lexington Herald-Leader* quickly condemned the proposal for going against KERA's philosophy that all schools can improve dramatically. Although the present study does not look at contextual effects on school performance *per se*, its findings may serve to moderate the concerns of parties on both sides of this controversy. No harm will come from adjusting scores to make the accountability procedures fairer. Still, these adjustments may not make much difference if the main focus of accountability remains on short-term improvements within schools and within districts.

The present findings have implications for the practice of ranking schools and districts. The results strongly suggest that the news media should refrain from ranking districts and schools based on their actual accountability index scores.⁵ Comparing districts (and probably schools) in this manner is biased and can seriously mislead the public and policy makers.⁶ It may also harm the self-esteem of educators, students, and parents in disadvantaged districts. In the interest of fairness and to ensure accurate perceptions, scores should be adjusted for contextual effects before rankings and comparisons are made. Since accountability scores may be sensitive to contextual effects at the higher grade levels, this caveat carries greater weight for ranking middle and high schools than elementary schools.

A crucial question remains which the findings of this study cannot answer: How will contextual effects influence the improvement potential of districts and schools over ten to twenty years? It is conceivable that districts might continue to improve their scores from one biennium to the next but that the long-term improvement of the disadvantaged districts would still lag far behind the advantaged districts. This possibility cannot be dismissed until a trend analysis is performed. Since the Kentucky Department of Education is substantially altering the accountability test and resetting the baseline in 1998-99, this very important issue will remain moot for some years while sufficient data are accumulated.

Finally, if it should happen that future research confirms the pattern of decline in the importance of contextual effects which was noted in the present study, this would offer dramatic confirmation that KERA is benefitting children's learning across the state. The reduction of contextual effects would mean that educational outcomes are becoming less sensitive to external sources of inequality, thereby achieving one of KERA's primary goals.

Conclusion

Repeated measures general linear modeling has been shown to be a useful tool to assess two principal ways in which contextual effects may bias high-stakes accountability scores. First, bias may affect test scores within educational units across successive biennia. Second, bias may occur when comparing test scores across educational units. These paired analyses are well suited for assessing if a high-stakes system of accountability testing contributes to greater equality of educational opportunity. The present study has not found evidence to fault Kentucky's method of within-district and within-school assessment. The influence of contextual effects on this procedure appear insubstantial for the most part. On the other hand, comparisons between districts (and presumably between schools) are quite sensitive to contextual effects. This is an important finding because in Kentucky the media and the public are prone to make such comparisons. The between-districts analysis also heightens the concern that substantial long-term improvement could be an unrealistic goal for disadvantaged school districts without the enhancement of local community resources. A longer period to monitor the influence of contextual effects on the trend of accountability scores is needed to clarify this issue.

Footnotes

⁵James Coleman (1988) introduced the concept "social capital" to embrace a set of factors that must be

HIGH-STAKES ACCOUNTABILITY AND CONTEXTUAL EFFECTS

considered when explaining student outcomes. Social capital includes norms, socialization processes, and social structures within the family and the community that enable children to succeed in school.

²Problems with twelfth graders' lack of incentive to take the tests brought about a change in test administration during the 1993-94 school year and afterwards. The cognitive tests were administered to eleventh graders instead of twelfth graders, but the writing portfolios were still required of twelfth graders.

³Other variables considered were the income gini coefficient for each school district, percent of the adult population with a high school degree, and racial heterogeneity of the local population. The first two variables were significantly related to the dependent variables in zero-order correlations but lost significance in multivariate analyses that included median household income. Racial heterogeneity does not vary greatly across Kentucky school districts. Thus, this factor was unrelated to the dependent variables even in zero-order correlations.

⁴Catterall, *et al.* (1998) argue persuasively that the increase in KIRIS scores has been influenced by familiarization with the test, teaching to the test, and numerous changes in the testing procedures. The results of the present analysis show that contextual effects probably do not need to be added to this list of confounding influences.

⁵See, for example, the report of school and district accountability index scores with rankings in Section B of the December 5, 1997 issue of the *Lexington Herald-Leader*.

⁶Guskey & Kifer (1990) reached a similar conclusion in their pre-KERA study.

References

- Alspaugh, J. W. (1992). Socioeconomic measures and achievement: Urban vs. rural. *Rural Educator*, 13, 2-7.
- Bickel, R., Weaver, S., Williams, T., & Lange, L. (1997). Opportunity, community, and teen pregnancy in an Appalachian state. *The Journal of Educational Research*, 90, 175-191.
- Catterall, J. S., Mehrens, W. A., Ryan, J. M., Flores, E. J., & Ruben, P. M. (1998, January). *Kentucky Instructional Results Information System: A Technical Review*. Louisville and Frankfort, KY: Coopers & Lybrand, L.L.P., and the Commonwealth of Kentucky Legislative Research Commission.
- Coleman, J. S. (1988). Social capital and the creation of human capital. *American Journal of Sociology*, 94 (Supplement), S95-S120.
- Coleman, J. S. (1990). *Equality and achievement in education*. Boulder, CO: Westview.
- Coleman, J. S., Campbell, E. Q., Hobsön, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Department of Education.
- Crane, J. (1991). The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *American Journal of Sociology*, 96, 1226-1259.
- DeYoung, A. J. (1991). Economic underdevelopment and its effects on formal schooling in southern Appalachia. *American Educational Research Journal*, 28, 297-315.
- Elmore, R. F., Abelman, C. H., & Fuhrman, S. H. (1996). The new accountability in state education reform: From process to performance." In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 65-98). Washington, DC: The Brookings Institution.
- Federal Reserve Bank of New York. (1998). Proceedings of a conference on excellence in education. *Economic Policy Review*, 4(1).
- Garner, C. L., & Raudenbush, S. W. (1991). Neighborhood effects on educational attainment: A multilevel analysis. *Sociology of Education*, 64, 251-262.
- Greenberg, E. J., & Teixeira, R. A. (1995). Nonmetro student achievement on par with metro. *Rural Development Perspectives*, 10, 17-23.
- Guskey, T. R. (Ed.). (1994). *High stakes performance assessment: Perspectives on Kentucky's educational reform*. Thousand Oaks, CA: Corwin Press.
- Guskey, T. R., & Kifer, E. W. (1990). Ranking school districts on the basis of statewide test results: Is it meaningful or misleading?" *Educational Measurement: Issues and Practice*, 9, 11-16.
- Hallinan, M. T. (1988). Equality of educational opportunity. *Annual Review of Sociology*, 14, 249-268.
- Hannaway, J., & Talbert, J. E. (1993). Bringing context into effective school research: Urban-suburban differences. *Educational Administration Quarterly*, 29, 164-186.
- Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, 19, 141-164.
- Khattari, N., Riley, K. W., & Kane, M. B. (1997). Students at risk in poor, rural areas: A review of the research. *Journal of Research in Rural Education*, 13, 79-100.
- Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petosky, M. D., Keselman, J. C., &

- Levin, J. R. (1998). Statistical practices of educational researchers: An Analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, 68, 350-386.
- Ladd, H. F. (1996). Introduction. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 1-19). Washington, DC: The Brookings Institution.
- Lee, V. E., Smith, J. B., & Croninger, R. G. (1997). How high school organization influences the equitable distribution of learning in mathematics and science. *Sociology of education*, 70, 120-150.
- Marino, T. W. (1995). Isolation, lack of role models blamed for low career expectations in rural areas. *Counseling Today*, 38(1), 12.
- Mathews, J. G., & Hackett, E. R. (1997, November). *The Alabama superintendent's report card: An analysis of local education agency characteristics and success at meeting state defined performance measures*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis, TN.
- Rodriguez, C., Jr. & Moore, N. B. (1995). Perceptions of pregnant/parenting teens: Reframing issues for an integrated approach to pregnancy problems. *Adolescence*, 30, 685-706.
- Schneider, B., & Coleman, J. S. (Eds.) (1993). *Parents, their children, and schools*. Boulder, CO: Westview.
- Smith, M. H., Beaulieu, L. J., & Israel, G. D. (1992). Effects of human capital and social capital on dropping out of high school in the south. *Journal of Research in Rural Education*, 8, 75-87.
- Stern, J. D. (1994). *The condition of rural schools*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Teachman, J. D., Paasch, K., & Carver, K. (1997). Social capital and the generation of human capital. *Social Forces*, 75, 1343-1359.
- Thompson, B. (1992). Two and one-half decades of leadership in measurement and evaluation. *Journal of Counseling & Development*, 70, 434-438.
- Zigarelli, M. A. (1996). An empirical test of conclusions from effective schools research. *The Journal of Educational Research*, 90, 103-109.

Developing an Identity of Third-Graders

Sharon L. Newbill

University of New Mexico

Andrea Clements

East Tennessee State University

This article begins to reconstruct the identity of third-grade school children through qualitative interpretation of their self-stories. We attempted to determine the nature of lived experience as written by the children, and from that, to find out what can be known about them. A sample of 188 autobiographies written by African-American, Caucasian, and Hispanic third-graders from eight urban, elementary schools was analyzed. The findings were categorized following qualitative coding procedures and collected into three psychological constructs which provided an interpretive scheme for understanding the responses: Self-Concept, Idealistic/Hypothetical Thinking, and Fears. The student identities, as reconstructed from the children's responses, are presented by gender and race and discussed with respect to normal psychological development and the available literature on minority identity structures. A pattern of responses emerged which crosscuts the psychological constructs, gender, and race. We define changes in the school culture needed to facilitate the development of well-adjusted children. Classroom teachers should encourage risk-taking behaviors, provide opportunities for personal recognition and empowerment, and actively promote personal safety in the school and community.

Introduction

This article begins to develop the identity of young, minority children by using autobiography. Our goal was to understand the meaning third-grade children attach to their lived experience. Our essential research question was "What is the nature of reality according to these youngsters, and from that, what can be known about them?" Answers to that question are a contribution to the literature on minority children's identity since few studies are published on this age group.

The use of a structured autobiography as a data collection tool reflects our concern with meaning and interpretation as preconditions for understanding (Denzin, 1989). Following a constructivist paradigm, we learned about third-graders through lives they have constructed in self-stories, which were then *re-constructed* by the researcher through interpretations of those writings (Guba & Lincoln, 1994). The self-stories are based in perceptions of reality. Yet, the meanings ascribed to that reality in order to make sense of it are constructed realities (Lincoln & Guba, 1985). Lincoln and Guba offer labels,

stereotypes, dictionaries, and theories as examples of constructed realities.

The analysis of the written constructions provided by the students was guided by a theory of hermeneutics, or the interpretation of the meaning contained in written expressions (Smith, 1993). The concern immediately arises about the "truth" of our representation. Denzin argues that, "autobiographies and biographies are only fictional statements with varying degrees of 'truth' about 'real' lives" (1989, p. 25). Herein lies a source of skepticism for those unfamiliar with qualitative research generally and interpretive biography specifically. We attempt to allay this skepticism initially by framing autobiography as personal narrative, then by acknowledging the challenges of re-construction.

The autobiography was chosen with the belief that a person is the only witness to his or her thinking and what people do is directed by what they think (Whitlock, 1986). Autobiography is purportedly a method which maintains the integrity of the writers perspective (Solas, 1992). This perspective serves to frame how students make sense of and interpret their world. Furthermore, we agree with Denzin in viewing self-stories as fictional, personal narratives:

Fiction is a narrative (story, account) which deals with real or imagined facts . . . Truth . . . refers to statements that are in agreement with facts . . . There are, then, true and false fictions; that is fictions that are in accord with facts and

Sharon L. Newbill is Sr. Research Scientist II in the Office of Evaluation, University of New Mexico. Andrea Clements is Associate Professor in the Department of Human Development and Learning at East Tennessee State University. Correspondence regarding this article should be sent to Sharon L. Newbill, Office of Evaluation, University of New Mexico, Albuquerque, NM 87131 or by e-mail at slnewbill@juno.com.

facticities [how facts are lived and experienced] as they are known or have been experienced, and fictions that distort or misrepresent these understandings. A truthful fiction (narrative) is faithful to facticities and facts. (1989, p. 23, italics in original)

We recognize the challenges confronting the researcher using self-stories as data. For example, we have the problems of a) whether the researchers are able to recognize the meanings contained in autobiographical writings and to interpret these meanings appropriately, b) whether the students have the ability to articulate their thoughts, c) the extent to which the student writes about his/her true self, d) whether the student is willing to share personal experiences, and e) the constraints of the form of representation (Newbill, 1996). Because our approach was interpretive, we will not be preoccupied with objective truth and validity (see Howard, Maerlender, Myers, & Curtin, 1992 to establish the validity of autobiographies). We were concerned with identifying common perceptions and meanings while maintaining the personal nature of lived experience. In order to achieve both data richness and integrity, we used consensus as the primary guarantor and creator of the "truth" of the constructed realities of the third-graders. We cross-validated individual self-stories by comparing them with accounts from other children in the sample of similar age, gender, and race. Through cross-validation, common or consensual categories of meaning emerged from the data that transcended gender and race (see Procedures). In conjunction with the literature on child psychology and identity structures we interpreted these consensual meanings as expressions of identity.

Rationale for Psychological Constructs

To facilitate explanation of the findings, the autobiographical questions are collected into three psychological constructs: Self-Concept; Idealistic/Hypothetical Thinking, and Fears. We think these constructs offer a credible and interpretive scaffold upon which we can begin to develop an understanding of third-grade identity. The constructs are well established in the psychological literature as explained below.

Self-Concept. Self-concept is as it sounds, one's concept of oneself. As early as 1902, self-concept was believed to be due largely to our interactions with others (Cooley, 1902). Self-concept is defined by Berk (1994) as "a set of beliefs about one's own characteristics" (p. 438). These characteristics include cultural identity and cognitive development. For example, in studies of self-concept in children from different cultures, noticeable differences in self-concept are found which reflect the values of that culture (Damon & Hart, 1988). Developmental changes in

self-concept have been shown through simple questioning such as asking the child to describe him or herself. Three to five year olds typically describe themselves by activities they perform, such as "I go to school" or "I play ball." By elementary school, children will begin to identify personal characteristics, such as eye color or hair color. Between ages 8 and 11 years, children will begin to mention personality traits such as whether they are outgoing or honest (Livesley & Bromley, 1973; Montmayor & Eisen, 1977). Because third-graders are typically eight or nine years old, one would expect to see a transition from physical descriptions to occasional mention of personality traits in writing about themselves.

One aspect of self-concept is self-esteem, which involves judgements about one's own worth. Children begin to make social comparisons around the time they begin elementary school (Erickson, 1950; Marsh, Craven, & Debus, 1991; Stipek & MacIver, 1989). These comparisons tend to cause a dip in self-esteem from kindergarten or first-grade until approximately third-grade. Beginning in the fourth-grade, self-esteem usually rises (Nottelmann, 1987; Wallace, Cunningham, & Del Monte, 1984). Child-rearing practices and other aspects of a child's natal culture greatly affect the child's self-esteem (Berk, 1994).

Our interest in self-concept is a challenging task given the very limited empirical research available on the development and expression of self-concept among minority children and adolescents. Available studies on minority children are either restricted to males or older age groups, or the findings are inextricably linked with racial identity or school success (e.g., Aires & Moorehead, 1989; Davidson, 1996; de Domanico, Crawford, & de Wolfe, 1994; Gordon, 1996; Mboya, 1994; Parham & Williams, 1993; Porter & Washington, 1993; Rice, Ruiz, & Padilla, 1974; Rotherman-Borus, 1989; Spencer, Cunningham, & Swanson, 1995; Thompson, 1992; Weiland & Coughlin, 1979). Minority cognitive styles, especially as a function of academic performance, have received some attention from researchers (Brookins, 1996; Helms & Parham, 1990). Yet, the findings from studies of cognitive styles are not informative to our study of self-concept.

Idealistic/Hypothetical Thinking. Hypothetical thinking is defined as "the ability to systematically generate and evaluate potential solutions to a problem" (Bukatko & Daehler, 1995, p. 217). The emphasis in our study on idealistic/hypothetical thinking came out of Piagetian Cognitive-Developmental theory (Piaget & Inhelder, 1969). This theory asserts that children progress through a series of cognitive stages in which they think in qualitatively different and more sophisticated ways as they get older. The children involved in this study fell in the age group included in Piaget's third stage, concrete operations. In this stage, children tend to be able to problem

solve, but only in familiar situations with information they understand. They are not able to think hypothetically about things they have not experienced, nor do they understand abstract information. The inability to understand abstract material can be shown by asking children in this age group to define an abstract word such as love or joy. Their answers typically will reflect experiences they have had or seen, and they will define the word by providing examples rather than a definition. Although children in the age group in this study are still quite concrete in their thinking, several of the questions in the structured autobiography required some degree of hypothetical thinking (e.g., "List the changes you would make in your home"; "A favorite wish is . . ."; "A future goal is . . ."). When children in the concrete operational stage (ages 7-11) are presented with items that are hypothetical in nature, they typically rely on quite realistic responses (i.e., things they have been taught, such as "you need math and science to succeed" and "reduce, reuse, recycle").

Questions in this construct also reflected an interest in the opportunity for idealistic thinking. The questions asked not only required hypothetical thinking but also led students to give idealistic responses (e.g., "changes you would make in your home"). It is known that children gradually become more idealistic until a peak in adolescence, returning to a somewhat more realistic outlook by early adulthood (Enright, Lapsley, & Shukla, 1979). These idealistic responses were sought at a time earlier than one would typically find idealistic thinking occurring spontaneously.

Fears. Fear in infancy and toddlerhood is elicited by loud noises or strange people. However, later in childhood fears are usually in response to more complex events such as school failure or rejection by peers (Morris & Kratchowill, 1983; Rutter & Garnezy, 1983). Fears expressed by elementary age children also tend to be related to more realistic rather than fictitious events. For example, a preschool child might be afraid of monsters under the bed, whereas an elementary child would more likely express fear over the possibility of burglars coming into the house. Fears actually decrease as children develop strategies for coping with scary situations (Berk, 1994), and the types of fears expressed and strategies for coping with them are closely related to a child's level of cognitive development (Bukatko & Daehler, 1995). The new found ability to use logic enables the child to understand how certain strategies (e.g., locking doors) can reduce the probability of the feared event. It is noteworthy that many of the fears expressed by children in this study were fears of things that were beyond their control, such as community violence. Their new ability to use logic does not help if there is nothing that they can do to lessen the threat of the feared event.

Methodology

Site and Sample

The children studied were enrolled in a large, urban school district in the Midwest that was under court-ordered desegregation. The segregation remedy involved the creation of magnet theme schools of choice rather than more involuntary methods, such as forced busing. Magnet theme schools offered very specific programs of study such as the arts, business, math/science, or computers. Parents and students chose from all schools in the district the school they wished for the students to attend. The theme of choice was supposed to draw (like a magnet) non-minority students from the suburbs and private schools back into district schools.

Regardless of the desegregation effort, the school district remained predominantly minority and of low socioeconomic conditions. During the school year the data were collected, approximately 37,000 students were enrolled in district schools. Of these, 68% were African-American, 25% Caucasian, 5% Hispanic, and 2% Asian-Pacific Islander. Seventy percent of all students were in the free/reduced lunch program. As compared to other "great cities," the school district was similar in percentage of minority students (approximately 75%, with a greater population of African-American than Hispanic as found in great city norms) and had a higher percentage of students enrolled in the free/reduced lunch program (70% in the district as compared to 55% of great city norms) (Clyde Hepner, personal communication, May, 1994). Furthermore, the attendance rate and the percentage of students scoring above the national norm in reading and math were lower in the school district as compared to the great city norms.

Table 1
Demographic Profile of Elementary Schools 1993-1994
Percentages of Total Enrollment by Race and SES

Sch	Enrollment N	n	Afr-Am	Cauc	Hisp	SES
M	320	64	78%	21%	<1%	77%
P	365	63	69%	29%	1%	79%
L	273	43	66%	32%	2%	74%
T	378	61	79%	21%	<1%	80%
A	295	42	73%	26%	<1%	80%
G	124	21	26%	37%	37%	60%
H	477	108	98%	1%	<1%	89%

Note. 'N' denotes total school enrollment and 'n' the number of third-graders. 'SES' is the percentage of students enrolled in the free or reduced lunch program.

Seven (of 45) elementary schools participated in the study. Six were magnet theme schools (three communi-

cations and writing (school M, P, and L), three visual and performing arts (school T, A, and G)) and one traditional school (school H). As presented in Table 1 and described below, some differences in student demographics were found among the schools. However, every school had a predominantly minority student population, and the majority of students were enrolled in the free/reduced lunch program schools.

The schools differed somewhat in student enrollment, minority composition, and socioeconomic status as measured by free/reduced lunch status. The latter is an imperfect measure, but the only one generally available. Not surprisingly, the traditional school (School H) had the largest population of students, greatest proportion of minority students, and highest percentage of students on free/reduced lunch. School G, a new school built as part of the desegregation effort and situated within a predominantly Hispanic neighborhood, differed markedly in demographic characteristics in that it was much smaller in enrollment and had a large proportion (37%) of Hispanic students. Also, the school offered a dual theme: a communications/writing theme and a foreign language (Spanish) theme. This school, although having a lower percentage of students enrolled in the free/reduced lunch program as compared to the other schools, was included in the study because it was the only source of a large number of Hispanic students. Excluding School G because of an imperfect socioeconomic measure would have excluded Hispanic representation in our study. So, while some between school differences exist, class differences among the races were negligible considering the sample as a whole.

The seven elementary schools were selected conveniently from all district schools. Our professional contacts within these schools greatly facilitated access to the teachers and students. Every third-grade student in the seven schools had the opportunity to complete the autobiography. The 400 returned and completed autobiographies were sorted by school. Then, using a proportionate random sampling of African-American and Caucasian students, 25 autobiographies were selected from each school. Given the small number of Hispanic students, all completed autobiographies written by them were included. The final sample ($n = 188$) is as proportionately representative of the racial diversity in the third-grade as was possible: 134 African-American (80 girls and 54 boys), 31 Caucasian (20 girls and 11 boys), and 23 Hispanic (13 girls and 10 boys) third-grade students.

Instrument

An autobiography was used to collect personal information about the third-graders. A suite of open-ended questions rather than a narrative was the preferred form of the autobiography given the youthfulness of the

population. Appropriate questions were generated from two sources: elementary school counselors and counseling and child psychology literature. The final form of the autobiography borrowed the presentation style and content segments from several counseling guides (Farnette, Forte, & Loss, 1989a, 1989b; Mannix, 1989; Palladino, 1989; Smith-Martenz, Cooper, & Leverte, 1977; Stanish, 1988). Children drew a self-portrait on the first page. The next pages presented a series of open-ended questions. The questions related to self-concept asked about their wishes, skills, accomplishments, what they are best at, a way they improve, changes they would make in themselves, and things they would not change. Questions requiring idealistic or hypothetical thinking involved future goals, changes in their home, changes in their world, a favorite wish, and a favorite daydream. The fear construct was defined by responses to one question regarding what they don't like to think about.

Procedures

The autobiography packets were distributed to the third-grade teachers in late January 1994 and were picked up at the end of April 1994. The three months given to complete the project allowed the teachers the opportunity to incorporate the autobiography into their regular instruction. Teachers were asked to have someone other than themselves present the packet to the students and, if possible, guide them through the project. Also, the teachers were asked not to read the autobiographies and to inform the students that he/she would not be reading their work. This was intended to give the students a sense of freedom in their responses. In most schools, the counselor introduced the packet to the students and was the primary facilitator. In only two schools did the teacher not have the staffing resources needed to assist in the task. In these schools, the teachers reported that they did not read the autobiographies, assured students each time of their use that they (the teachers) would not be reading them, and openly put them away after the class finished work for the day. Lastly, the researchers were not present in the schools at any time during the writing and were not involved with the students in any way.

The teachers assisted with this project by providing identifying data on each autobiography, specifically student number, gender, racial code, school, and grade-level. A large part of the success in obtaining so many completed autobiographies is attributed directly to the cooperation of the teachers and their principals.

The autobiographies were written by approximately 1000 students in the third-, fifth-, sixth-, eighth-, ninth-, tenth-, and twelfth- grades. Those written by the third-graders are reported here, with findings presented by gender and race.

Data analysis included both qualitative and quantitative techniques. Quantitative methods were used to obtain the frequency of responses, which served to identify common categories. Next, qualitative coding procedures were used to further reduce the responses to themes (Marshall & Rossman, 1994; Strauss & Corbin, 1990). For example, in response to the question about their favorite wish many students wrote something about being rich and famous. The response was entered originally as the student wrote it, and, since many students gave a similar response, the category "rich/famous" was created for that particular question. In the discussion below, the abbreviated autobiographical question is italicized to enable the reader to distinguish between original intent, student construction, and researcher re-construction.

Results

Racial Identity and Gender Structures

African American. Considering first the construct self-concept, African-American children did not differentiate between *skills* and *accomplishments*. They felt equally skilled and accomplished in sports and in academics. Feelings of skill and accomplishment among females were located in fine arts and, for males, in traditional sports. For example, the children wrote of feeling accomplished from "being good at perfect attendance" (female), "join b-ball" (male), "I won a trophy in football" (male), and "be good" (female). They felt skilled at "football" (male), "dancing" (female), and "math" (male and female).

The self-esteem of African-American children seemed positive as they liked themselves. For instance, in response to being asked to "*List things that you would not change*," the most frequent response of both males and females was either "myself" or "my family." Probing further through a more specific follow-up question, "*What would you change about yourself*," the responses differed by gender. Males would improve their behavior (e.g., "not fight" or "change my addtude [sic]") whereas females would change their physical appearance (e.g., "my hair" or "face"). When asked *what it would take to improve*, the children recognized that changes required effort, such as "practice" and "do it often."

What children *wished to become* and how they saw the *future* required idealistic or hypothetical thinking. African-Americans, as with all children in the study, *wished to be rich and famous*. Also, shared by all children regardless of race or gender, was a *future goal* of being a celebrity. The males wanted to become well known in sports (e.g., "to play in the NBA"), and the females, through the arts (e.g., "an artist that makes a lot of things"). The African-American children *day-dreamed of*

going fun places, often the local amusement park, or being rich and famous (e.g., "I was in Worlds of fun" (male and female), "I day dream I was rich" (male), or "to become like Tina Turner and to graduate" (female). The African-American children had specific *changes to their home*, which most often was to have a room of their own.

One hypothetical question, that elicited thoughts about *changes in the world*, yielded responses striking in accord with their fears. African-American children wanted less violence in the world (e.g., "change killing people" (male and female), "no shooting" (male), or "fighting" (female)). Asking, "*I don't like to think about . . .*" captured the children's fears. "Death" was the overwhelming response. These constructs reveal an imagination filled with violence and its aftermath.

Caucasian. Regarding self-concept, the responses of the sample of Caucasian children differed very little from their African-American classmates. Children of both races, regardless of gender, felt *skilled* in some area of sports or academics. Male children of both races gained a sense of *accomplishment* through traditional sports. Unlike African-American females, however, Caucasian girls felt a sense of accomplishment through academics rather than fine arts ("becoming a good reader" "reading science").

The level of self-esteem may be somewhat higher in African-American children than Caucasian as demonstrated by not wanting to change themselves when asked what they *would not change*. Many Caucasian children wrote they would not change "my family" or "my parents" rather than themselves as written by the African-Americans. Interestingly, the *changes in themselves* Caucasian children perceived were identical to those of the African-Americans: males would change their behavior or attitude ("not have a bad temper" or "acting up") and females selected some aspect of their physical appearance ("be prettier").

Expressions within the construct of idealistic/hypothetical thinking were common among all races. Regardless of race, children wished to be rich; to become famous through sports (males) or the arts (females); they day-dreamed of going places ("to go to Worlds of Fun") or improving themselves ("to grow up"); and having a room of one's own. However, the *change for the world* envisioned by Caucasian children differed markedly from their African-American classmates. Caucasian children focused on a cleaner environment ("recycle more" or "no trash") rather than less violence as perceived by the African-American children. However, regardless of the degree of violence in their lives, like their African-American schoolmates, the greatest fear of Caucasian children was death.

Hispanic. Hispanic children differentiated among the questions targeting self-concept more than did their African-American and Caucasian classmates. For example, Hispanic children, regardless of gender, felt *most skilled* in sports ("basketball" or "running"), yet gained the greatest sense of *accomplishment* through academics ("getting my name on the board" or "getting all A+ on my grade card"). Unlike the African-American and Caucasian children's self-concept, Hispanics did not include themselves in the list of *things they would not change*. Rather, they felt changes in themselves were needed in three areas -- for males, better behavior ("not akting [sic] bad"); for females, better looking ("my figer [sic]"); for both, being smarter ("smert" [sic]). The only thing Hispanic children would not change was their family. Family before self was a differentiating signature in the Hispanic identity structure.

In the construct of idealistic/hypothetical thinking, Hispanic children *daydreamed* most frequently about their family. Interestingly, instead of personal space as found in Caucasian and African-American children, the Hispanics wanted more power at home. For example, the *changes at home* most often mentioned were having a voice in house rules ("my rules").

As stated above, favorite wish and future goals of Hispanic children did not differ from the other racial groups. Hispanics, like the African-American children, wanted less violence in the world ("no killing"). However, the prevalent fear of Hispanic third-graders was school rather than death. The fear of school among Hispanic children can be plausibly located in low academic self-concept and unsettling school experiences. The data show a self-concept of feeling less capable academically and of needing to be smarter and better behaved in school.

Summary

Self-concept

Common to most of the children, regardless of race or gender, was a need to change themselves. However, the wording of one item ("I would change about myself...") may have led children, who may not have otherwise been concerned with changing anything about themselves, to think of something that needed changing. The perceived change differed by gender, but not race. Males felt a need to improve their behavior and females to alter their physical appearance. Our findings of the greater focus on physical beauty and the academic-orientation of females when compared to males agrees with the literature in that, beginning in preschool, females are more conforming to adults and peers than are males (Block, 1976; Macoby & Jacklin, 1974).

Self-concept did not differ remarkably between African-American and Caucasian children. Both races

identified with sports or academics, with more males than females associating with sports, and vice versa. Since males are more physical than females (Cummings, Iannotti, & Zahn-Waxler, 1985), it is not surprising that they identified more strongly with sports than did females. Self-esteem is tied to academic, physical, and social competence between the ages of 6-10 years (Erikson, 1950). This was seen in the responses of these children. Self-esteem may drop as children begin to compare themselves with others in early elementary grades (Berk, 1994) but increases with grade level (Franco, 1983; Larned & Muller, 1979).

The Hispanic children were distinguished easily from the other racial groups in the area of self-concept, and their responses generally paralleled the female children. In comparison, both Hispanics and females placed a greater emphasis on family and reflected a lower self-concept. The data on self-concept indicated that females and Hispanic children wanted to be prettier (females) or smarter (Hispanic) and gained a sense of pride through academics, yet felt most competent in sports. On the other hand, males and African-American and Caucasian children felt equally accomplished and skilled in sports and in academics. Neither female nor Hispanic children would change their family. Males, African-Americans, and Caucasians (to a lesser extent) would not change themselves primarily and considered their family second.

The above findings may be interpreted as revealing a lower self-concept among females and among Hispanic children. This interpretation is supported by research on multiethnic children and adolescents that demonstrate lower measures of self-concept among Mexican American children (Franco, 1983; Knight, Kagan, Nelson, & Gumbiner, 1978; Widaman, MacMillan, Hemsley, Little, & Balow, 1992) and young girls (Gilligan, 1991, 1982; Rogers, 1993; Rogers, Brown, & Tappan, 1994).

Idealistic/Hypothetical Thinking

As discussed above, children are just beginning to think logically and realistically in third-grade. When they made attempts at idealistic thinking, they typically gave either familiar situations they have experienced (e.g., going to amusement parks, etc.), or fairly farfetched "storybook-type" responses (e.g., wanting all the money in the world, etc.). According to Piaget and Inhelder (1969), these children still think concretely and are bound in their thinking by their own experiences. Our findings supported those developmental theories.

Most children in the study wished to become rich and famous. Fame, as defined through a future goal, was viewed differently by males and females. Females saw a more diffuse and less definitive future than males, which could reflect either a lack of female role models or less of a tendency to seek and emulate role models. Males

THIRD-GRADE IDENTITY

seemed to have more distinct roles to which to aspire. For example, a future goal was being a sports figure, policeman, fireman, or detective. Responses of most females were more diffuse in writing generally of being a good person.

Going someplace, as seemingly mundane as the mall or park, was a frequent daydream among most children, especially females. These findings could indicate that female children felt more confined and constrained than did males. There is some evidence that this might be true in that males are allowed to roam the neighborhood and to range farther from home unsupervised than are females (Newson & Newson, 1976).

Hispanic children, on the other hand, daydreamed about trips to see family in distant places rather than fun outings to local entertainment venues. Recall the finding from self-concept in that, comparatively, Hispanic children seemed more secure in their family than in themselves. It is fitting to include here the finding that when thinking about changes in their home, the Hispanic children wanted more voice in the home rather than a separate space away from them as expressed by the other children. The findings agree with Marcell (1994) by showing that Hispanic students stressed the importance of family. A greater emphasis on family expressed consistently by the Hispanic students may be related to the maintenance of cultural traditions and the contact with extended family (Harrison, Wilson, Pine, Chan, & Buriel, 1990).

Children did not focus on interpersonal issues, which is normal for this age. However, some mention was made of interpersonal concerns, with males more so than females wanting fewer fights and arguments at home. Typically, males are more negatively affected by household discord and divorce than are females. Changes males reported wanting to make in their world typically involved reducing violence. This concern was prevalent among all students, except the Caucasians, who placed the environment above violence. Shakoor and Chalmer (1991) found in their survey of 1,035 African-American adolescents from seven Chicago area schools that three-fourths had witnessed someone being shot, stabbed, robbed, or killed. The authors termed this witnessing "co-victimization," and such exposure could be a contributor to the findings here since all African-American children in the sample reside in the innermost part of a city ranked among the highest in violent crime and drug trafficking at the time of this study.

Fears

Fear of more realistic things begins around seven to eight years of age and becomes well developed between 9 and 12 years (DeAngelis, 1991; Morris & Kratchowill, 1983). In the preschool and early elementary school years

children's fears focus on unrealistic images such as the "bogey man," but by third-grade mention of specifics such as war, violence, AIDS, and so forth are common. The children in this study had reached the age at which they could think about what they did not want to think about, which suggests that they had developed some degree of control over their thinking, or metacognition. The fact that these students understood this question indicates a developmental milestone.

Fears again distinguished the Hispanic from African-American and Caucasian children. School was the predominant fear of Hispanic children whereas death was the pervasive fear of the other races. A realistic understanding of the permanence and universality of death generally does not occur until approximately age seven (Berk, 1998), so these children had just begun to understand death. Additionally, many of them, by virtue of their living conditions, had been exposed to death, particularly through violence, as well. The fear of school among Hispanics complements the findings from the self-concept construct in which Hispanic children felt inadequate in school.

The results can be summarized using a figure. Figure 1 presents a summary of the self-concept, idealistic/hypothetical thinking, and fears of African-American, Caucasian, and Hispanic male and female third-grade children.

Discussion

The identity of third-graders as revealed in this study increases our understanding of these children and suggests areas educators can explore to better facilitate the learning process. A critical area is the classroom culture. It is significant to note that, although most Hispanic students felt accomplishment through academics, far fewer felt they were best at it. Findings suggest that Hispanics did not perceive school to be as welcoming a place as did African-American and Caucasian students. All children this age were aware of the importance of improving through trying harder and most students associated a sense of accomplishment with academics. Therefore, only gentle persuasion would be needed in classroom in which children from diverse backgrounds feel comfortable taking risks and extending themselves cognitively and emotionally. Educators in multiethnic settings may need to examine their pedagogy with the aim of enabling a more culturally relevant learning environment (see Osborne, 1996). Since students already seem to understand the importance of academic success, future efforts might be most effective if they assume that understanding, inform the children that they are aware of their understanding of the importance of schooling, and hold high expectations for the children.

Self-Concept						
	African-American		Caucasian		Hispanic	
	Male	Female	Male	Female	Male	Female
Best at	Sports/Academics	Sports/Academics	Sports/Academics	Academics	Sports	Sports
Feel accomplishment	Sports/Academics	Sports/Academics	Sports/Academics	Academics	Academics	Academics
Would NOT change	Self and/or Family	Self and/or Family	Self and/or Family	Self and/or Family	Family	Family
Change self	Behavior	Appearance	Behavior	Appearance	Behavior	Appearance
					Smarter	Smarter
Idealistic/Hypothetical Thinking						
	African-American		Caucasian		Hispanic	
	Male	Female	Male	Female	Male	Female
Wish to be	All children regardless of gender or race wished to be RICH and FAMOUS					
Future goal	All children regardless of gender or race wished to become a CELEBRITY					
Daydream	Local fun spot	Local fun spot	Local fun spot	Local fun spot	Family	Family
Change in home	Room of own	Room of own	Room of own	Room of own	Some authority	Some authority
Change in world	Less violence	Less violence	Cleaner environment	Cleaner environment	Less violence	Less violence
Fears						
	African-American		Caucasian		Hispanic	
	Male	Female	Male	Female	Male	Female
Don't like to think about	Death	Death	Death	Death	School	School

Figure 1. A summary of the findings regarding the self-concept, idealistic/hypothetical thinking, and fears of third-graders.

Another area of application for these findings is in the area of personal recognition or empowerment. Children this age want to be noticed. Receiving recognition in school gives them a sense of accomplishment. The responses to the autobiographies suggest that trips to a local spot (such as Discovery Zone, Worlds of Fun, Cool Crest, a mall, or a park) would be an influential way of recognizing and rewarding third-graders. An even less costly but effective means of gaining recognition would be permitting a space of their own in the classroom. The simple need for these third-graders to have a room of their own could be addressed in the school by allotting a particular space for each child. Perhaps, having lockers at this age would be a way of meeting this need. Interestingly, the need for their own space did not seem to be as prevalent among Hispanic children as it was among the others. For Hispanics, a voice or sense of ownership in the classroom, such as in the making of classroom rules, could be an effective form of recognition.

Finally, violence was a significant concern in these children's lives. There is a need to constructively address

violence in the community and how best to deal with it emotionally and psychologically. But the children's feelings about threats or potential violence should be addressed as well. Maslow (1943), who theorized about motivation, indicated that students who are concerned about safety issues are not able to concern themselves with higher level needs which would include academic pursuits. Only when children are helped to feel less afraid will they be able to focus more on school.

This research began to examine the differences in experience, skills, expectations, and life styles children bring into the classroom. Further research using a fifth-grade sample contemporary to these third-graders both in time and location will allow us to develop a comparative understanding of how cultural, gender, and grade-level differences influence a child's educational experience. An awareness of the behavioral implications of such differences in a school setting may provide teachers with tools to provide a form of cultural therapy (Spindler & Spindler, 1994) for these students which may facilitate their learning experiences.

THIRD-GRADE IDENTITY

References

- Aires, E., & Moorehead, K. (1989). The importance of ethnicity in the development of identity of Black adolescents. *Psychological Reports*, 65, 75-82.
- Berk, L. E. (1998). *Development through the lifespan*. Boston, MA: Allyn and Bacon.
- Berk, L. E. (1994). *Child development* (3rd ed.). Boston, MA: Allyn and Bacon.
- Block, J. H. (1976). Issue, problems, and pitfalls in assessing sex differences: A critical review of "The psychology of sex differences." *Merrill-Palmer Quarterly*, 22, 283-308.
- Brookins, C. C. (1996). Exploring psychosocial task resolution and self-concept among African-American adolescents. *Perceptual and Motor Skills*, 82, 803-810.
- Bukatko, D., & Daehler, M. W. (1995). *Child development: A thematic approach* (2nd ed.). Boston, MA: Houghton Mifflin Company.
- Cooley, C. H. (1902). *Human nature and the social order*. New York: Scribner.
- Cummings, E. M., Iannotti, R. J., & Zahn-Waxler, C. (1985). Influence of conflict between adults on the emotions and aggression of young children. *Developmental Psychology*, 21, 495-507.
- Damon, W., & Hart, D. (1988). *Self-understanding in childhood and adolescence*. New York: Cambridge University Press.
- Davidson, A. L. (1996). *Making and molding identity in schools: Student narratives on race, gender, and academic achievement*. Albany, NY: State University of New York Press.
- DeAngelis, T. (1991, March). Psychologists take calls from kids about war. *APA Monitor*, 22(11), 8.
- de Domanico, Y. B., Crawford, I., & De Wolfe, A. S. (1994). Ethnic identity and self-concept in Mexican-American adolescents: Is bicultural identity related to stress or better adjustment? *Child & Youth Care Forum*, 23, 197-207.
- Denzin, N. K. (1989). *Interpretive biography. Qualitative Research Methods*, Volume 17. Newbury Park, CA: Sage Publication.
- Enright, R. D., Lapsley, D. K., & Shukla, D. (1979). Adolescent egocentrism in early and late adolescence. *Adolescence*, 14, 687-695.
- Erikson, E. H. (1950). *Childhood and society*. New York: Norton.
- Farnette, C., Forte, I., & Loss, B. (1989a). *I've got me and I'm glad*. Nashville, TN: Incentive Publications, Inc.
- Farnette, C., Forte, I., & Loss, B. (1989b). *People need each other*. Nashville, TN: Incentive Publications, Inc.
- Franco, J. N. (1983). A developmental analysis of self-concept in Mexican American and Anglo school children. *Hispanic Journal of Behavioral Sciences*, 5, 207-218.
- Gilligan, C. (1991). Teaching Shakespeare's sister: Note from the underground of female adolescence. *Women's Studies Quarterly* 1 & 2, 31-51.
- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.
- Gordon, K. A. (1996). Resilient Hispanic youths' self-concept and motivational patterns. *Hispanic Journal of Behavioral Sciences*, 18, 63-73.
- Guba, G. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.) *Handbook of Qualitative Research* (pp.105-117). Thousand Oaks, CA: Sage Publications.
- Harrison, A. O., Wilson, M. N., Pine, C. J., Chan, S. Q., & Buriel, R. (1990). Family ecological of ethnic minority children. *Child Development*, 61, 347-362.
- Helms, J. E., & Parham, T. A. (1990). The relationship between Black racial identity and attitudes and cognitive styles. In J. E. Helms (Ed.) *Black and white racial identity: Theory, research and practice* (pp. 119-131). New York: Greenwood Press.
- Howard, G. S., Maerlender, A. C., Myers, P. R., & Curtin, T. D. (1992). In stories we trust: Studies of the validity of autobiographies. *Journal of Counseling Psychology*, 39, 398-405.
- Knight, G. P., Kagan, S., Nelson, W., & Gumbiner, J. (1978). Acculturation of second- and third-generation Mexican American children: Field independence, locus of control, self-esteem, and school achievement. *Journal of Cross-Cultural Psychology*, 9, 87-97.
- Larned, D. T., & Muller, D. (1979). Development of self-concept in Mexican American and Anglo students. *Hispanic Journal of Behavioral Sciences*, 1, 279-285.
- Lincoln, Y. S., & Guba, G. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.
- Livesley, W. J., & Bromley, D. B. (1973). *Person perception in childhood and adolescence*. London: Wiley.
- Macoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
- Mannix, D. (1989). *Be a better student: Lessons and worksheets for teaching behavior management in grades 4-9*. New York: Simon & Schuster.
- Marcell, A. V. (1994). Understanding ethnicity, identity formation, and risk behavior among adolescents of Mexican descent. *Journal of School Health*, 64, 323-327.

- Marsh, H. W., Craven, R. G., & Debus, R. (1991). Self-concept of young children 5 to 8 years of age: Measurement and multidimensional structure. *Journal of Educational Psychology, 83*, 377-392.
- Marshall, C., & Rossman, G. B. (1994). *Designing qualitative research*. Newbury Park, CA: Sage Publications.
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*, 370-396.
- Mboya, M. M. (1994). Multidimensional self-concept: Relations with age and gender among African adolescents. *Educational Psychology, 14*, 307-321.
- Montmayer, R., & Eisen, M. (1977). The development of self-conceptions from childhood to adolescence. *Developmental Psychology, 13*, 314-319.
- Morris, R. & Kratchowill, T. (1983). *Treating children's fears and phobias: A behavioral approach*. Elmsford, NY: Pergamon.
- Newbill, S. L. (1996, March-April). *Autobiographies as a qualitative research tool*. Paper presented at the meeting of the Eastern Educational Research Association, Cambridge, MA.
- Newson, J., & Newson, E. (1976). *Seven years old in the home environment*. London: Allen & Unwin.
- Nottelmann, E. D. (1987). Competence and self-esteem during transition from childhood to adolescence. *Developmental Psychology, 23*, 441-450.
- Osborne, A. B. (1996). Practice into theory into practice: Culturally relevant pedagogy for students we have marginalized and normalized. *Anthropology & Education Quarterly, 27*, 285-314.
- Palladino, C. D. (1989). *Developing self-esteem: A positive guide for personal success*. Canada: Crisp Publications, Inc.
- Parham, T. A., & Williams, P. T. (1993). The relationship of demographic and background factors to racial identity attitudes. *Journal of Black Psychology, 19*, 7-24.
- Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. London: Routledge & Kegan Paul.
- Porter, J. R., & Washington, R. E. (1993). Minority identity and self-esteem. *Annual Review of Sociology, 19*, 139-161.
- Rice, A. S., Ruiz, R. A., & Padilla, A. M. (1974). Person perception, self-identity, and ethnic group preference in Anglo, Black, and Chicano preschool and third-grade children. *Journal of Cross-Cultural Psychology, 5*, 100-108.
- Rogers, A. (1993). Voice, play, and a practice of ordinary courage in girls' and women's lives. *Harvard Educational Review, 63*, 265-295.
- Rogers, A., Brown, L., & Tappan, M. (1994). Interpreting loss in ego development in girls: Regression or resistance? In R. Josselson & A. Lieblich (Eds.) *The narrative study of lives* (pp. 18-43). Newbury Park, CA: Sage.
- Rotherman-Borus, M. J. (1989). Ethnic differences in adolescents' identity status and associated behavior problems. *Journal of Adolescence, 12*, 361-374.
- Rutter, M., & Garmezy, N. (1983). Developmental psychopathology. In E. M. Hetherington (Ed.), *Handbook of child psychology: Vol. IV. Socialization, personality, and social development*. New York: Wiley.
- Shakoor, B. H., & Chalmer, D. (1991). Co-victimization of African-American children who witness violence: Effects on cognitive, emotional, and behavioral development. *Journal of the National Medical Association, 83*, 233-238.
- Smith, J. K. (1993). Hermeneutics and qualitative inquiry. In D. J. Flinders & G. E. Mills (Eds.), *Theory and concepts in qualitative research: Perspectives from the field* (pp. 183-200). Columbia University: Teachers College Press.
- Smith-Martenz, A., Cooper, J., & Leverte, M. (1977). *Giving kids a piece of the action*. Doylestown, PA: Mar*Co products, Inc.
- Solas, J. (1992). Investigating teacher and student thinking about the process of teaching and learning using autobiography and repertory grid. *Review of Educational Research, 62*, 205-225.
- Spencer, M. B., Cunningham, M., & Swanson, D. P. (1995). Identity as coping: Adolescent African-American males' adaptive responses to high-risk environments. In H. W. Harris, H. C. Bluc, & E. E. H. Griffith (Eds.), *Racial and ethnic identity: Psychological development and creative expression* (pp. 31-52). New York: Routledge.
- Spindler, G., & Spindler, L., (Eds.) (1994). *Pathways to cultural awareness: Cultural therapy with teachers and students*. Newbury Park, CA: Corwin Press.
- Stanish, B. (1988). *The giving book: Creative classroom approaches to caring, valuing and cooperating*. Carthage, IL: Good Apple, Inc.
- Stipek, D., & MacIver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development, 60*, 531-538.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory, procedures, and techniques*. Newbury Park, CA: Sage Publications.
- Thompson, V. L. S. (1992) A multifaceted approach to the conceptualization of African-American identification. *Journal of Black Studies, 23*, 75-85.
- Wallace, J. R., Cunningham, T. F., & Del Monte, V. (1984). Change and stability in self-esteem between late childhood and early adolescence. *Journal of Early Adolescence, 4*, 253-257.

THIRD-GRADE IDENTITY

- Weiland, A., & Coughlin, R. (1979). Self-identification and preferences: A comparison of White and Mexican-American first and third graders. *Journal of Cross-Cultural Psychology, 10*, 356-365.
- Whittrock, M. (1986). Students' thought processes. In M. Whitrock (Ed.), *Handbook on research on teaching* (3rd ed., pp.297-314). New York: Macmillan.
- Widaman, K. F., MacMillan, D. L., Hemsley, R. E., Little, T. D., & Balow, I. H. (1992). Differences in adolescents' self-concept as a function of academic level, ethnicity, and gender. *American Journal of Mental Retardation, 96*, 387-404.

Measures of Work Culture in High and Low Performance Schools

Darlene Y. Bruner
Valdosta State University

Bobbie Jean Greenlee
University of South Florida

This study sought to determine if significant differences existed in the work cultures of schools with high student achievement scores and schools with low student achievement scores. Quantitative findings indicated differences in work culture between Low and High Performance Schools. The School Work Culture Profile provided information regarding levels of involvement of staff found in both groups and yielded a statistically significant difference ($p < 0.0002$) between them. The Education Quality Benchmark System provided information about work patterns in relation to Quality philosophy and discriminated between the Low and High Performance Schools ($p < 0.03$). The results add to the literature regarding the importance of culture in educational organizations in relation to student performance.

The primary purpose of this study was to establish the relationship between measures of work culture and student achievement in schools. The hypothesis for this study was that schools with higher student achievement indicators would have more highly developed quality work cultures than schools with low student performance. Work culture defines the way individuals respond to one another and the expectations of the work that is to be done. Highly developed work cultures exhibit more interdependency, more involvement of all role groups, and more customer focus -- all of which are manifested in quality organizations (Bonstingl, 1992; Snyder, 1994, Snyder, Acker-Hocevar, & Snyder, 2000). The assumption made by the researchers was that the dynamics of work culture influence performance (Deal & Peterson, 1991; Lewis, 1993; Ott, 1989; Snyder & Anderson, 1986).

As social institutions, such as schools, become more responsive to changing needs, bureaucratic structures and processes of formal organizations will be replaced with transformed work cultures (Snyder, 1994). Accordingly, many of the deep-rooted traditions of schooling should be replaced with dynamic and energetic work cultures. Total

Quality Management (TQM), a management philosophy that institutionalizes planned and continuous improvement, fosters a change in the organization's core values and integration of quality principles into the work culture (Hill, 1995; Sallis, 1992). Within this framework, work becomes value driven with a shared vision and a collaborative work culture.

Transformation of the school work culture, based on a literature review by Snyder (1994), is enhanced by using a framework for becoming a quality system. The major themes Snyder found on quality were: a) customer satisfaction is the focus for work, b) visionary leadership drives the change process, c) systems thinking enables the organization to respond quickly to customer needs, d) strategic planning is essential for improving quality, e) continuous training in collaboration and in the use of information systems empowers workers to meet challenges routinely, and f) continual improvement toward quality becomes a way of life in the organization.

The culture is the overarching feature of a quality organization, which provides the context for work. Transforming work structures from the control emphasis of bureaucratic systems to responsive patterns found in quality systems requires attention to development of the work culture (Deal & Peterson, 1991; Lewis, 1993; Snyder, 1994). As educators seek to improve student performance through continual school improvement efforts, benchmarking where schools are in relation to systems thinking and quality concepts may prove beneficial to those efforts. Schools can learn to expose, examine, and transform old patterns to induce effectual change.

Darlene Y. Bruner is Assistant Professor at Valdosta State University, and Bobbie J. Greenlee is Visiting Assistant Professor at University of South Florida. We appreciate the comments of Dr. Anthony Onwuegbuzie, Valdosta State University, on this article. Correspondence concerning this article should be addressed to Darlene Bruner, Valdosta State University, Department of Educational Leadership, 1500 North Patterson Street, Valdosta, GA 31698. Electronic mail may be sent: dbruner@valdosta.edu.

Educational institutions, as organizations, lack helpful systems and models to use in transforming systems of work. Identification of the characteristics of work cultures found in schools with low student achievement based on norm-referenced standardized tests (Low Performance Schools) and schools with high student achievement based on the same measures (High Performance Schools) may assist schools in transforming their systems of work to meet the needs of students. There is a need to establish whether a difference exists in the work cultures found in Low and High Performance Schools and to find useful and helpful instruments to measure and benchmark school progress.

The Education Quality Benchmark System (EQBS) (Snyder & Acker-Hocevar, 1994) was designed to help schools chart the direction of change initiatives. It offers schools the opportunity to measure their performance against global standards for quality. The EQBS, a relatively new diagnostic tool, was designed to provide a quality framework to assist educators in developing more responsive work cultures over time.

The School Work Culture Profile (SWCP) (Snyder, 1988) provides descriptive information on levels of staff involvement in the work culture. It was designed to obtain a measure of professional involvement in the school's work patterns that have developed over time. The SWCP is based on the generalization that a school's work culture either hinders or supports educational productivity and quality.

Studies have found compelling evidence for improving school culture. Fyans and Maehr (1990) found that students are more motivated to learn in schools with strong cultures. Thacker and McInerney (1992) found significant effects on student achievement in elementary schools when the school focused on creating a new mission statement, created goals based on outcomes for students, aligned the curriculum to meet those goals, and provided professional growth opportunities. School culture also correlates with teacher productivity and satisfaction. Cheng (1993) found that school cultures with strong commitment, shared participation, intimacy, and strong leadership had more motivated teachers.

Methods

In Florida, public reports contain common data elements to allow statewide reporting of each school's progress toward the state's eight education goals. Summaries of information about school achievement, learning environment (or process), and student characteristics are published as Vital Signs reports to the public (Florida Department of Education, 1993). The academic indicators

are the school scores on nationally normed standardized tests for Reading Comprehension, Math Concepts and Applications, and the Florida Writes test. The criteria for schools, at the time of this study, was for 33% or more of the students tested to be at or above the 50th percentile on each of the three academic indicators over a two year period. These three school scores over two consecutive years yield the six academic indicators by which schools are judged.

Participants

Known groups of six elementary schools were studied to help identify overall work culture trends and patterns. These schools reflected high and low student achievement based on Florida Vital Signs criterion. Three of the schools studied were High Performance Schools and met the criteria (33% of the students at the 50th percentile or above) for 5 or more of the 6 academic indicators. Three schools studied met the criteria for only one or two of the six academic indicators and were considered to be Low Performance Schools. The state did not officially deem schools as high or low performing; however, they report the data in descending order, with schools meeting all the academic indicators at the top of the list. The designation of low and high performance is made by the researchers based on the information supplied by the state of Florida.

Respondents in the study were asked to provide information about themselves, their participation in school and district work groups, and their perception of schools in general. This information was analyzed for features and patterns of the Low and High Performance Schools. Survey data from the Education Quality Benchmark System (EQBS) and the School Work Culture Profile (SWCP) were used with 144 teachers and administrators from six elementary schools. Participation was voluntary, and of those surveyed, 69% participated.

Demographic information about those schools characterized as High Performance Schools indicated that, overall, they had higher student and staff attendance and lower levels of poverty and minority students (Table 1). However, one school in the High Performance Schools group had a relatively high proportion of students eligible for free and reduced lunch (60%) and a high percentage of minority students (41%). This High Performance School had been receiving Title I service through the school year prior to this study. Title I provides additional money from the federal government to enhance personnel and material resources for schools with high poverty rates. All Low Performance Schools were Title I schools with higher than the District average eligible for free or reduced lunch.

MEASURES OF WORK CULTURE

Table 1
Demographics of Schools in Sample

School	Enrollment	Free/Reduce Lunch	Student Minority	Exceptional Staff Attendance	Student Attendance	Enrollment	Gifted
HIGH A	760	28%	22%	95%	96%	9%	4%
HIGH B	817	34%	20%	95%	97%	5%	2%
HIGH C	440	60%	41%	96%	97%	7%	3%
LOW D	395	79%	27%	93%	97%	8%	1%
LOW E	469	78%	46%	93%	95%	13%	1%
LOW F	504	64%	45%	94%	95%	16%	0%

Data Analysis

Instruments

School Work Culture Profile. The SWCP contains 60 five-point Likert-type items ranging from 1 = Strongly Disagree to 5 = Strongly Agree. The items pertain to school work practices and are organized into four subscales: a) Organizational Planning, b) Professional Development, c) Program Development, and d) School Assessments. Each subscale has 15 items such that scores range from 15 to 75, with scores indicating levels of staff involvement in the school's work patterns. These scores can be summed to obtain a total score that ranges from 60 to 300, measuring the construct of work culture within an organization or system. According to Snyder (1988), school's work culture refers to the combined practices of planning, organizing, delivering, monitoring, and assessing as related to teaching and learning. The combined score provides information regarding the work patterns found in High and Low Performance Schools.

The SWCP has been subject to several validation and reliability studies. These studies provide high reliability estimates and evidence for construct validity for the SWCP (Johnson, Snyder & Johnson, 1992; Parkinson, 1990). For the present sample, reliability coefficients, as measured by Cronbach's Coefficient Alpha were as follows: Low Performance Schools (.96), High Performance Schools (.99), and the total sample (.98). The magnitude of these reliability estimates of the total scale is consistent with the alphas reported in previous reliability studies (.95 and above).

The SWCP provides information on the strength of the school's interaction system in the planning, development, delivery and assessment of programs and services. The SWCP was developed based upon a systems model of collaborative school management, and assesses the level of staff involvement (Snyder & Anderson, 1986).

Education Quality Benchmark System. According to its authors, the EQBS is a diagnostic tool for educators to use in assessing the interdependency of work systems in schools (Snyder & Acker-Hocevar, 1994). The EQBS can provide benchmarks for examining change over time. This instrument consists of items measuring Performance Areas that enable schools to assess the quality of their services and programs. The EQBS has been shown to generate reliable scores, and evidence of content and concurrent validity has been established (Acker-Hocevar, 1994; Greenlee, 1997; Snyder, 1997).

The EQBS is a 48-item instrument designed around the dimensions of work within the Quality Performance System Model (Snyder & Acker-Hocevar, 1994). The dimensions called Performance Areas include: a) Visionary Leadership, b) Strategic Planning, c) Systems Thinking and Action, d) Information Systems, e) Human Resource Development, f) Quality Services, and g) Customer Success and Satisfaction. Outcomes are identified for each Performance Area, with each having four indicators across phases of organizational development. Respondents choose the indicator that best describes their current practices. The response options were weighted by the numerical values 4, 3, 2, 1, and 0, respectively assigned to the response categories Transformed, Transition, Awareness, Bureaucratic, and Do Not Know. Scores of the EQBS range from 0 to 192 with high scores indicating "Quality-like" work patterns in the organization.

Reliability estimates in split half analysis were computed and corrected using the Spearman-Brown Prophecy Formula for low performing schools (0.90), high performing schools (0.96), and for the total sample (0.94). Cronbach's Alpha reliability estimates, which assess the inter-item consistency, were high. The alpha coefficient for the total sample was 0.97, while the alpha for low and high performing groups was 0.95 and 0.98, respectively.

The EQBS identifies the interdependencies of the dimensions of work and its impact on customer success and satisfaction (Snyder & Acker-Hocevar, 1994). The EQBS is a diagnostic tool to better align the systems of work by: continuously innovating programs to enhance services, transforming structures to a responsiveness orientation, and working interdependently to achieve new purposes.

Results

Measures of central tendency show High Performance Schools had a higher mean (120.95) than Low Performance Schools (111.75) on the EQBS. High Performance Schools had a higher mean (244.15) than the Low Performance Schools (231.72) on the SWCP (Table 2).

Table 2
SWCP and EQBS Means and Standard Deviation
by Performance Group

Schools	SWCP		EQBS	
	M	SD	M	SD
High Performance	244.15	43.44	120.95	48.22
Low Performance	231.72	28.88	111.75	31.92

A Chi-square test comparing the scores on each subscale of the SWCP between Low and High Performance Schools (Table 3) indicated statistically significant differences for the subscales of Planning ($p < 0.03$), Program Development ($p < 0.003$) and Staff Development ($p < 0.02$). The scores for the Assessment subscale did not differ significantly between the two groups. There was a significant difference ($p < 0.0002$; $df = 2$) between the High and Low Performance Schools on the SWCP.

Table 3
SWCP Subscale Means by Performance Group

SWCP Subscales	Low Schools		High Schools		Effect Size
	M	SD	M	SD	
Planning	57.54	7.26	61.29	10.87	.52
Program Development	60.25	7.47	63.10	11.76	.38
Staff Development	58.59	7.87	60.98	11.97	.30
Assessment	55.35	8.28	58.79	10.24	.42

Chi-square tests comparing the scores on each Performance Area of the EQBS between Low and High Performance Schools (Table 4) indicate statistically significant differences for Visionary Leadership ($p < 0.0002$), Information Systems ($p < 0.002$), Human Resource Development ($p < 0.03$), and Customer Success and Satisfaction ($p < 0.03$). The scores for Systems Thinking and Action,

Strategic Planning, and Quality Services did not show significant differences between the two groups. There was a statistical significance ($p < 0.03$; $df = 3$) between the High Performance and Low Performance Schools on the EQBS.

Table 4
EQBS Performance Area Value Label Means and Effect Sizes
for High and Low Performing Schools

EQBS Subscales	Low Schools		High Schools		Effect Size
	M	SD	M	SD	
Visionary Leadership	2.49	.69	2.70	1.04	.30
Strategic Planning	2.35	.84	2.52	1.00	.20
Systems Thinking and Action	2.35	.74	2.61	.91	.34
Information Systems	2.26	.80	2.58	1.06	.40
Human Resource Development	2.36	.81	2.65	1.01	.36
Quality Services	2.32	.77	2.07	.81	-.32
Customer Success & Satisfaction	2.39	.82	2.72	.97	.41

Note: Means are presented by value label 4, 3, 2, 1, and 0, respectively assigned to the response categories Transformed, Transition, Awareness, Bureaucratic, and Don't Know.

The known groups were restricted based on the schools' performance on student achievement indicators, and presumed differences in work cultures. Results from the two instruments revealed that differences did exist between the work cultures in Low and High Performance Schools.

The statistical significance of the data may not be informative about the practical importance of the findings. It has been suggested that an index of the practical significance can be derived from an effect size estimate (Cohen, 1977; Harris & Rosenthal, 1985; Rosenthal, 1983). The effect size represents the magnitude of the difference between the means of the two groups. It is figured by taking the difference of the means and then dividing by the standard deviation for the control group. Using Cohen's (1988) criteria, that a correlation of 0.5 is large, 0.3 is moderate, and 0.1 is small, all effect sizes corresponding to the differences in means between the high and low performance schools were moderate to large. The effect sizes suggest notable differences between the two groups.

The effect sizes for the mean scores on the subscales of each of the two instruments used in this study provide some idea about how schools with high or low student achievement may be influenced by their work cultures. A positive effect size would indicate that work culture appears to be more developed and more "quality-like" in schools with high student achievement.

MEASURES OF WORK CULTURE

Discussion

The selection of known groups was based on the assumption that schools with high student achievement indicators would have more developed work cultures than schools with low student achievement indicators. Using two instruments that measure work cultures in distinct ways we have found that schools grouped based on student achievement have significantly different work cultures.

Schools in this study with higher student achievement were reported to have different and more collaborative work cultures than schools with lower student achievement. The EQBS indicators depict information systems, networks, performance standards, and a customer orientation, whereas the SWCP examines goal setting, work group planning, clinical supervision, instructional management, resource development, and assessment. Both the SWCP and the EQBS identified the differences in the work patterns of the schools.

Information from the indicators in the EQBS and the SWCP provide a report of schools with quality-like work cultures. In schools with highly developed work cultures, improvement processes seem to be centered on bettering student performance through specific training of teachers. Principals communicate the vision and the direction of the school by words and deeds to all whose cooperation may be needed to accomplish school goals. Innovations and contributions of teachers are recognized. The environment supports design and redesign of programs and experimenting with options to meet the needs of the students. These schools study their performance routinely and acquire and share new knowledge.

The purpose of this study was to determine how quality principles, as a means of developing school work culture, influence student achievement. The positive moderate effect sizes would indicate achievement appeared to be higher for students in quality-like, well-developed work cultures. In all subdomains of the instruments there were moderate positive effect sizes with the exception of the Quality Services Performance Area of the EQBS. The negative effect size found in Quality Services may indicate that the particular interventions in that Performance Area may decrease student achievement or may not discern differences in student achievement.

Performance, to a degree depends on the interaction with other people in the organization. Sharing and collegiality, empowerment, and leadership are characteristics commonly found in schools attempting to generate and sustain cultural improvement (Jakicic, 1994; Simpson, 1990). Harris (1985) points out that culture cannot be

installed or prescribed; it emerges from appropriate practices and behaviors. A quality work culture is the result of a common vision that guides change in the school and adds value to the services (Snyder, Wolf, Acker-Hocevar, 1995).

The distinctiveness of the school culture is contextualized by participants, who, each in their own way, state and defend attitudes, beliefs, and theories about the purpose of education, the nature of teaching, the role of the teachers, how children learn, and how school settings should be organized (Sarason, 1995). Culture also affects views about the nature of work and the worker, about standards of satisfactory performance and productivity, and worker education and training. Culture develops slowly, affecting our identity, behavior, and performance (Harris, 1985). Efforts to change schools without addressing the school's culture are at risk from the beginning.

The success of the quality movement in industry to change the culture of organizations and improve performance has prompted consideration of how to apply quality principles to educational organizations. Significant implications from this study indicate that schools with higher student achievement have many quality principles and systems in place. The direction and quality of work in an organization are stimulated by culture (Snyder, 1988, Snyder, Acker-Hocevar, & Snyder, 2000). Paying attention to the culture of schools appears to affect performance.

Ultimately, quality fosters change in the school's culture and in the teaching and learning process. The integration of quality principles in the organization's core values promotes the pursuit of quality through the process of continual improvement (Acker-Hocevar, 1996; Wolverton, 1993). A quality culture develops systemwide responses to the changing needs of students that may result in student success. Educational leaders wanting to improve their schools for the benefit of the students should focus on the quality principles that enhance the culture of the organization. Organizational excellence emerges, or fails to emerge, based on the culture the organization creates.

References

- Acker-Hocevar, M. (1994). *The education quality system: A content validation*. Tampa, FL: A dissertation at the University of South Florida.
- Acker-Hocevar, M. (1996). Conceptual models, choices, and benchmarks for building quality work cultures. *NASSP Bulletin*, 80(576), 78-86.

- Bonstingl, J. (1992). The quality revolution in education. *Educational Leadership*, 50(3), 4-9.
- Cheng, Y. C. (1993). Profiles of organizational culture and effective schools. *School Effectiveness and School Improvement*, 4(2), 85-110.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: John Wiley Publishing.
- Deal, T. E. & Peterson, K. D. (1991). *The principal's role in shaping school culture*. Washington, D.C.: U.S. Department of Education.
- Florida Department of Education. (1993). *Blueprint 2000: A system of school improvement and accountability for 1994-95 school improvement plans*. Tallahassee, FL: Florida Commission on Educational Reform and Accountability.
- Fyans, M.G. & Maehr, M. L. (1990). *School culture, student ethnicity, and motivation*. Urbana, IL: The National Center for School Leadership. ED 327947.
- Greenlee, B. (1997). *A study of the internal consistency reliability and criterion related validity of the Education Quality Benchmark System*. A dissertation at University of South Florida at Tampa.
- Harris, M. J. & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, 97, 363-386.
- Harris, P. R. (1985). *Management in transition*. San Francisco, CA: Jossey-Bass, Inc.
- Hill, S. (1995). From quality circles to total quality management. In Wilkinson, A. & Willmont, H. (Eds.) *Making quality critical: New perspectives on organizational change*. London: Routledge.
- Jakicic, C. (1994). Taking small steps to promote collaboration. *Journal of Staff Development*, 15, 16-18.
- Johnson, W. L., Snyder, K. J., & Johnson, A. M. (1992). Developing instruments for educational administration: An exposition using primary and second-order factor analysis. *National Forum for Applied Educational Research Journal*, 6(1), 3-11.
- Lewis, J. (1993). *Implementing total quality in education to produce great schools: Transforming the American school system*. New York: National Center to Save Our Schools.
- Ott, J. S. (1989). *The organizational culture perspective*. Chicago, IL: The Dorsey Press.
- Parkinson, A. (1990). *An examination of the reliability and factor structure of the school work culture profile instrument*. Tampa, FL: A dissertation at the University of South Florida.
- Rosenthal, R. (1983). Assessing the statistical and social importance of the effects of psychotherapy. *Journal of Consulting and Clinical Psychology*, 51, 4-13.
- Sallis, E. (1992). *Total quality management in education*. Philadelphia, PA: Kogan Page.
- Sarason, S. B. (1995). *School change: The personal development of a point of view*. New York: Teachers College Press.
- Simpson, G. (1990). Keeping it alive: Elements of school culture that sustain innovation. *Educational Leadership*, 47, 34-37.
- Snyder, K. J. (1988). *School work culture profile*. Tampa, FL: Managing Productive Schools Training Programs.
- Snyder, K. J. (1994). The quality revolution in schools: A Florida partnership venture. *International Journal of Educational Reform*, 3(3), 279-289.
- Snyder, K. M. (1997). *A construct validation and reliability estimation of the educational quality benchmark system*. Tampa, FL: A dissertation at the University of South Florida.
- Snyder, K. J. & Acker-Hocevar, M. (1994). *Education quality benchmark system*. Tampa, FL: School Management Institute, College of Education at University of South Florida.
- Snyder, K. J., Acker-Hocevar, M., & Snyder, K. (2000). *Living on the edge of chaos: Leading schools into the global age*. Milwaukee, WI: Quality Press.
- Snyder, K. J. & Anderson, R. H. (1986). *Managing productive schools: Toward an ecology*. Orlando, FL: Academic Press.
- Snyder, K. J., Wolf, K. M., & Acker-Hocevar, M. (1995, September). *Changing schools to quality work cultures: Issues and dilemmas*. Paper presented at the British Educational Management and Administration Society, Oxford, England..
- Thacker, J. L. & McInerney. (1992). Changing academic culture to improve student achievement in the elementary schools. *ERS Spectrum*, 10(4), 18-23.
- Wolverton, M. (1993). *Total quality management*. ERIC Digest, 73. Eugene, OR: ERIC Clearinghouse on Educational Management.

JOURNAL SUBSCRIPTION FORM

This form can be used to subscribe to *RESEARCH IN THE SCHOOLS* without becoming a member of the Mid-South Educational Research Association. It can be used by individuals and institutions.

Please enter a subscription to *RESEARCH IN THE SCHOOLS* for:

Name: _____

Institution: _____

Address: _____

		COST	
Individual Subscription (\$25 per year)	Number of years	_____	_____
Institutional Subscription (\$30 per year)	Number of years	_____	_____
Foreign Surcharge (\$25 per year, applies to both individual and institutional subscriptions)	Number of years	_____	_____
Back issues for Volumes 1 - 6 (\$30 per Volume)	Number of Volumes	_____	_____
TOTAL COST:			

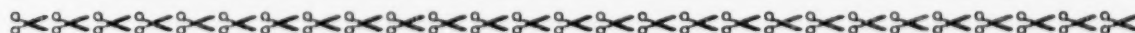
MAKE CHECKS PAYABLE TO MSERA
SEND FORM AND CHECK TO:

Dr. James E. McLean, Co-Editor
RESEARCH IN THE SCHOOLS
Warf-Pickle Hall, Room 418
Eastern Tennessee State University
Box 70685
Johnson City, TN 37614-0685

MID-SOUTH EDUCATIONAL RESEARCH ASSOCIATION

The Mid-South Educational Research Association (MSERA) was founded in order to encourage quality educational research in the Mid-South and to promote the application of quality educational research in schools. Members of MSERA share interests in educational research, development, and evaluation. While most members are from institutions of higher education, many others represent state departments of education, public and private agencies, and public school systems. Graduate students comprise a significant portion of the membership. A majority of MSERA members are from the six states represented by the organization, but others are from many other states and several foreign countries. The MSERA is the largest regional educational research organization in the country.

The organization provides several services for its members. The annual meeting, held every November, offers many formal and informal opportunities for professional development through special training courses, sharing of research findings, and programmatic interests with colleagues. Members receive a subscription to *RESEARCH IN THE SCHOOLS* and the *Mid-South Educational Researcher*. The MSERA also provides recognition and cash rewards for its outstanding paper, an outstanding dissertation, and professional service.



MSERA Membership/Renewal Form

(Please print or type)

Name _____

Organization _____

Address _____

Telephone Work: _____

Home: _____

Fax: _____

e-mail: _____

Amount Enclosed:	MSERA 1999 Membership (\$25 professional, \$15 student)	\$ _____
	MSER Foundation Contribution	\$ _____
	TOTAL	\$ _____

Make check out to MSERA and mail to:

Dr. Robert Calvery
Flippin School District
P. O. Box 239
Flippin, Arkansas 72634

RESEARCH IN THE SCHOOLS
Mid-South Educational Research Association
and the University of Alabama at Birmingham
901 South 13th Street, Room 233
Birmingham, AL 35294-1250

BULK RATE
U.S. POSTAGE
PAID
PERMIT NO. 1256
BIRMINGHAM, AL

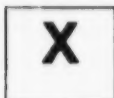


U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)



NOTICE

Reproduction Basis



This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.



This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").